Production of Online Word of Mouth: Peer Effects and the Moderating Role of User Characteristics*

Yang Wang,[1] Paulo Goes,[1] Zaiyan Wei,[2] and Daniel Zeng[1]

[1]Eller College of Management, University of Arizona, Tucson, AZ 85721
Emails: yangwang@email.arizona.edu, pgoes@eller.arizona.edu, zeng@eller.arizona.edu

[2]Krannert School of Management, Purdue University, West Lafayette, IN 47907. Email: zaiyan@purdue.edu

December, 2016

We study peer effects on users' contributions to an emerging form of online word of mouth—product reviews. Provided by either consumers or third-party professionals, online reviews are closely correlated with consumers' purchasing decisions and product sales. Individuals have incentives of free riding and maximizing social capital when providing feedbacks online. We leverage a "natural experiment," which leads to an exogenous expansion in the users population of a major online reviews platform, to better understand the trade-off between the two conflicting incentives. Our empirical findings are mainly two-fold. First, we find that a larger population of audience and peer review writers, an immediate consequence of the exogenous shock, causally leads to more reviews posted, higher and more diverse ratings assigned, and reviews of higher quality by the users. In addition, we find that these effects are moderated by user characteristics including activeness, expertise, and popularity. These results have implications for platforms that rely on user contributions, as well as for business in the management of their online product or service feedbacks.

Key words: online reviews, ratings, user-generated content, social media analytics, natural experiment

---

* We would like to thank Wei Chen, Mohit Tawarmalani, Tiemen Woutersen, Bin Zhang, and participants at the 7th Annual Conference on Internet Commerce and Innovation at Northwestern University, 2016 INFORMS Annual Meeting, and 2017 ASSA/AEA Annual Meeting for helpful comments and suggestions. Usual claims apply.

1. Introduction

Online product reviews, provided by either consumers or third-party professionals, are shown to be closely correlated with consumers' purchasing decisions and product sales (Dellarocas 2003, Chevalier and Mayzlin 2006, Liu 2006, Duan et al. 2008, Forman et al. 2008, Zhu and Zhang 2010). The majority of online retailing platforms, including Amazon.com (Mudambi and Schuff 2010) and eBay.com (Li and Hitt 2008), allow consumers to post product feedbacks and opinions. Platforms specialized in facilitating third-party product reviews, such as Yelp.com and Douban.com, are also emerging in recent years (Chen and Xie 2005). According to an online survey in 2014,[1] only 10% consumers do not take any form of online reviews. Among those who pay attention to product reviews, 88% trust them as much as personal recommendations.

With the increased importance of online consumer feedbacks, companies and their business managers now pay more attention to this emerging form of word of mouth (Gu and Ye 2014, Abrahams et al. 2015, Chen et al. 2016, Sun and Xu 2016). Although the consequences are well understood, individuals' incentives to produce online reviews in the first place are not equally treated. We are among the first few studies investigating the antecedents to online product reviews (Goes et al. 2014, Wang et al. 2015, Khern-am-nuai and Kannan 2016). Regardless of the various motivations, a fundamental factor governing individuals' contributions is peer effects. Similar to other forms of user-generated content, an individual's reference groups, online friends/followers/audience, have significant influence on his or her review posting behaviors (Zhang and Zhu 2011, Peng et al. 2016, Wei and Xiao 2016). In particular, contributions to online reviews have close correlations with the number of followers (Goes et al. 2014), and online friendship aggravates review similarities (Wang et al. 2015). In sharp contrast, based on theories predicting individuals' contributions to online public goods in a network setting, we use a "natural experiment" on a major third-party platform to identify the size effects of one's audience and peer groups.

Online reviews are public goods on the Internet (Duan et al. 2008). Anyone cannot be effectively excluded from "consuming" online reviews, and an individual's "consumption" of reviews does not

[1] The Local Consumer Review Survey conduct consumer studies on a yearly basis. The 2014 briefing can be found at `http://searchengineland.com/88-consumers-trust-online-reviews-much-personal-recommendations-195803`.

crowd out others' access. With this in mind, our theoretical foundations are three-fold. First, about the *volume* of contributions, it is not hard to see that there are conflicting incentives for individuals to contribute to this public good. With a larger size of the reference group, one has the incentive to free ride others' contributions and post less reviews (Olson 1965, Chamberlin 1974). On the flip side, individuals may be encouraged to post more out of the incentives to add to social capital and maximize social benefits (Harsanyi 1980, Andreoni 1989, 2007, Zhang and Zhu 2011). Second, social image considerations (Andreoni and Bernheim 2009, Lacetera and Macis 2010) dictate that individuals exhibit more prosocial behaviors in a larger network, therefore tend to assign higher *ratings* for products under review. Last but not least, with similar prosocial considerations, an individual will pay more attention to the *quality* of reviews facing a larger group of audience or peer review writers.

We test these predictions empirically in this study. In particular, we ask which incentive—free riding or prosocial considerations—dictates their private provision of the public good—product reviews, and how their activities are affected by the size of the reference groups. Specifically, do more peer review writers or a larger population of audience necessarily cause a review writer to devote more efforts and post more product reviews? In addition to the volume effects, does an increase in the network size make a review writer assign higher ratings in their product reviews? Is the helpfulness or the quality of reviews, evaluated by review readers, attenuated by an enlarged population of audience and peer reviewers?

We seek answers to these questions by studying Douban.com, the largest platform of third-party reviews in China. Launched on March 6, 2005, Douban.com specializes in providing a platform for consumers and professionals to write reviews for books, movies, TV shows, and music. The platform witnessed a sudden expansion in the number of registered users starting from August 1, 2009, due to the introduction of a web application, "Douban Reading," on the largest social networking platform in China, the QZone of Tencent.com.[2] The application granted Tencent users Douban accounts,

[2] QZone can be found at `http://qzone.qq.com/`. The Wikipedia page for QZone is `http://en.wikipedia.org/wiki/Qzone`. According to "We Are Social" reports, the number of active users on QZone were more than 644 million as of May 2014, making it the second largest online social network worldwide, only smaller than Facebook.com.

and direct access to all Douban book reviews. This exogenous shock, the unexpected merge of two large-scale and influential networks, provides us a unique opportunity to study individuals' incentives to contribute to online reviews in an exogenously enlarged network.

We collected our data using the API service provided by Douban.com in November 2014. We start with a random sample of 119 Douban groups, and focus on $24,374$ users in these groups, who joined the platform before the exogenous shock. We obtain all relevant information from their webpages using a web data crawler. The main study period is between July 4, 2009 and August 28, 2009, which is four weeks before and four weeks after the exogenous shock. There were no policy changes other than the introduction of Douban Reading during this period. The identification of peer effects (Manski 1993) comes from a unique feature of the exogenous shock. Specifically, Douban Reading users gained direct access to only the book reviews within the app, while it did not allow access to other sections unless the users went to the corresponding pages on Douban.com. This particular arrangement divides Douban reviews into two groups, book reviews being the treatment group and all other reviews the control group. With this "natural experiment," we conduct a quasi-experimental study, the differences-in-differences analysis (Card and Krueger 1994, Athey and Imbens 2006), to identify the effects of the policy change (a sudden increase in the population of registered users) on individual users' contributions to product reviews.

We first notice that the number of registered users on Douban.com was more than doubled within a month (skyrocketed from about 3.8 million to more than 8 million) following the introduction of Douban Reading on Tencent QZone. Not just the group of peer review writers exploded, individual reviews also drew larger groups of audience. Using the differences-in-differences method, we find that the exogenous shock led to about 0.1% increase in the volume of comments a review received on average. Not surprisingly, the merge of the two networks indeed led to a larger population of review writers as well as audience on Douban.com.

Our empirical findings are mainly two-fold. We first establish the average effects of an enlarged peer group, in the sense that we do not explore the differential effects by user characteristics in this set of results. Interestingly, in our main sample, the number of book reviews rose from 310 to 322. While in sharp contrast, other types of reviews (movies, TV shows, or music) all dropped by around 20%. Motivated by this finding from summary statistics, we conduct more formal tests by running regression-adjusted analysis. We find that the exogenous merge between Douban and Tencent caused a user to post more and longer reviews, assign higher and more diverse ratings in their reviews, and post reviews of higher quality. Quantitatively, an average user increased the number of reviews by more than 0.6%, and the textual lengths by over 4.2%. The rating assigned in each review rose by about 0.03 in absolute values (the range of rating is from 1 to 5 in integers). As a measure of a review's quality, the percentage of helpfulness votes (%) increased by about 0.27. All these findings are statistically significant. We make sure the robustness of these findings by taking care of a few concerns including the sampled users, the sampling period, users' experience, and the construction of the control group.

More interestingly, the second set of findings concerns the moderating role of user characteristics in the peer effects identified above. Arguably, we would expect differential impacts of the exogenous shock by a user's engagement with the platform (or activeness), his or her expertise in reviews for products, and the popularity. We first find that a user who was more engaged with Douban.com, measured by a larger number of early reviews (prior to our study period) or the number of followees (those whom a user follows), was less affected by the merging event in the volume, valence (the rating), and quality of reviews. A user who focused in writing reviews in other categories (movies, TV shows, or music) was also less affected, but in sharp contrast a book review enthusiast was encouraged to contribute even more reviews, assign even higher ratings, and post reviews of even higher quality. However, the moderating effects of one's popularity, measured by the number of followers, were negative but not statistically significant.

The study first contributes to the literature on online reviews (Dellarocas 2003, Chevalier and Mayzlin 2006, Goes et al. 2014, Lee et al. 2015, Wang et al. 2015, Chen et al. 2016). We start filling the gap in this literature by exploring individuals' incentives to post reviews, and the causal effects of the reference group size. Goes et al. (2014) and Wang et al. (2015) are the two closest studies. Goes et al. (2014) study how a user's current review-posting behavior is affected by his or her popularity in the last period, and Wang et al. (2015) show that online friendship aggravates review similarities. In sharp contrast, we explore a review writer's incentives to post product reviews in a larger network, and identify the size effects of the reference group using an exogenous expansion in the population of cohort reviewers and audience. Second, we add to the long-standing literature on the private provision of public goods (Olson 1965, Chamberlin 1974, Andreoni 1989). Our results provide another piece of empirical evidence that supports the conjecture of maximizing social benefits being a generic part of the incentives for an economic agent (Zhang and Zhu 2011). Last but not least, we also contribute to the broader literature on users' contributions in online communities (Gu et al. 2007, Ransbotham et al. 2012) by illustrating how community members respond to the growth of the community.

The paper is structured as follows. Section 2 summarizes related literature and develops our research hypotheses. We introduce our research context, Douban.com and the exogenous shock, in Section 3. We describe the datasets and present summary statistics in Section 4. We then provide detailed discussions of our empirical strategies in Section 5.1. We report our empirical findings in Section 5.2 and 5.3. Section 6 discusses and concludes.

2. Related Literature and Research Hypotheses

Online reviews can be considered as public goods (Duan et al. 2008). Specifically, product reviews are *non-excludable* as the review pages are generally public on the Internet and accessible to any users. In addition, a user's access to reviews does not preclude others' ability to access (*non-rivalry*). In the private provision of public goods, contributors gain utilities from not only contributions by the whole group but also personal contributions (Andreoni 1989).

Early literature on public goods provision hinges on the assumption that individuals gain utilities from only the volume of contributions by the whole group (Olson 1965, Chamberlin 1974). They predict that individuals' contribution level decreases with the group size—"free riding" phenomenon. However, pointed out by more recent studies (Andreoni 2007), these models fail to explain the extensive giving behavior observed in field and laboratory studies. Therefore, recent behavioral research extends the assumption, and formulate the idea that individuals also gain utilities from personal contributions (Harsanyi 1980, Andreoni 1989, Duncan 2004, Andreoni 2007). A fundamental conjecture of the new theories is that when the group size is sufficiently large, the relative importance of social benefits dictates individuals' incentives to contribute. Thus with sufficiently large reference group, individuals' contribution levels increase with the group size. In contrast to early public goods theory, empirical findings generally support the social benefit hypothesis (Andreoni 2007, Zhang and Zhu 2011).

In our specific context, theory suggests that review writers gain utilities (personal and social benefits) from not only the total stock of reviews on the platform, but also their own contributions. As the theory predicts, users will find it more beneficial to contribute more reviews with a larger population of audience. Thus our first set of hypotheses reflects the conjecture of social effects dominating users' incentives to produce online reviews.

Hypothesis 1a. (Volume I) *A larger number of members in a network cause individuals to write and post a larger number of product reviews.*

Hypothesis 1b. (Volume II) *A larger number of members in a network cause individuals to write and post on average longer product reviews.*

In addition to the *volume* effects, the size of audience population has potential impacts on the *valence* of reviews—review ratings. On third-party platforms including Douban.com, the review writers can choose an integer or a number of stars as their overall ratings of products. Goes et al. (2014) find a negative correlation between the average rating and the number of followers up to the previous period. They attribute the finding mainly to the fact that readers or consumers find negative evaluations more informative and professional (Ofir and Simonson 2001, Bateman et al. 2011). However, the effects of peer-group size can be biased due to the endogeneity of the followers count. For example, users gain expertise and followers simultaneously on the platform as they write more reviews, which may lead to attenuation of the true network size effects.

We draw on the social image theory that usually concludes that social image concerns are a primary motivator of prosocial behavior, such as the private provision of public goods (Andreoni and Bernheim 2009, Lacetera and Macis 2010). The argument is that people care about fairness and how they are perceived by the public (Andreoni and Bernheim 2009). As an example, Lacetera and Macis (2010) find that, in a series of field experiments, blood donors significantly increase the frequency of their donations only if the events are publicly announced in local newspapers, and they are awarded in public ceremonies. By analogy, review writers also care about how they are perceived by their audience when they produce reviews and assign ratings to products. The theory suggests that, with a larger population of review readers—the audience, review writers are more likely to assign higher ratings out of altruism (Andreoni and Bernheim 2009) and social image considerations (Lacetera and Macis 2010).

Hypothesis 2. (Valence) *A larger number of reference group members increase the overall valence of reviews (the ratings) assigned by the review writers.*

The dispersion of product ratings also has influence on product sales. A product with more diversified but lower average rating has higher demand subsequently (Sun 2012). Goes et al. (2014) hypothesize that due to the special structure of the rating system (with lower and upper bounds), the change in average rating leads to higher variance of ratings. Following the same logic, we also hypothesize that a larger population of audience causes users to assign more diverse ratings.

Hypothesis 3. (Valence Dispersion) *The variance of review ratings that individuals assign is higher with a larger population of audience and peer review writers.*

The "quality" of online reviews helps consumers better evaluate the quality of products. This feature has been studied in the literature and shown to be critical for consumers' information distraction and final product sales (Mudambi and Schuff 2010, Cao et al. 2011). Factors that influence the quality or helpfulness of reviews include reviewers' expertise, textual characteristics, and timing of posting online reviews in terms of the product lifecycle (Mudambi and Schuff 2010, Cao et al. 2011). That being said, the quality of product reviews may also be influenced by the population size of one's audience.

Previous studies establish that larger communities are likely to maintain postings of lower quality (Gu et al. 2007). Gu et al. (2007) theorize that as a community becomes larger, it has more incentive to provide bundled information due to economies of scale, rather than the quality of postings. Essentially, the community faces a tradeoff between fixed costs (bundled information that does not change with the community size) and marginal costs (the quality of each posting). In contrast, users of third-party review communities do not face a similar tradeoff. Instead, they are weighing the benefits and costs from posting reviews of high quality. On one hand, by providing more helpful reviews (more information, or more useful as a purchasing guide), review writers contribute more benefits to the community. The aforementioned theories of prosocial behavior and social image (Andreoni 1989, 2007, Andreoni and Bernheim 2009) both imply that users will improve the quality of reviews with more audience and peer reviewers. On the flip side, the costs of posting reviews clearly increase with the quality or helpfulness. With a larger population of audience, users spend more time and efforts in composing more reviews, thus will pay much less attention to the quality of their reviews. Therefore, it is not clearcut *a priori* in which direction a larger group of review readers and cohort review writers will change the quality or helpfulness of reviews. We therefore propose the following competing hypotheses:

Hypothesis 4a. (Quality I) *The "quality" or helpfulness of product reviews improves with the number of reference group members.*

Hypothesis 4b. (Quality II) *The "quality" or helpfulness of product reviews does not improve with the number of reference group members.*

3. Douban.com and Douban Reading

Launched on March 6, 2005, Douban.com was originally advertised as a platform providing social networking service. In contrast to other networking sites such as Facebook.com and Twitter.com, the friendship on Douban.com is mainly based on users' common interests in books, movies, TV shows, or music. The network structure is identical to that on Twitter.com. Specifically, a user can send a request to follow another user. Upon approved, a following relationship is established with the request sender being the follower and the receiver the followee. The follower will be able to see all activities by the followee, including all his or her reviews ever posted. The followee's all subsequent updates will show up in the follower's news feed thereafter.

More important, in addition to forming connections on the platform, all Douban users are allowed to rate and write reviews for all products in the platform's collection. The platform creates and maintains webpages for all its collected products.[3] Douban users can read and write product reviews, make comments below the reviews, and chat with other users in real time on the product pages. These are in sharp contrast to main activities in other social networking platforms, where sharing life events, news, or articles with friends are among the top activities. By the end of 2014, the number of registered Douban users had surpassed 100 million, with over 200 million monthly unique visitors and average daily page views (PV) beyond 210 million. On the other side of the market, Douban.com has over 17 million books, 320 million movies or TV shows, and 1 million songs in storage.[4] Douban.com is now the largest platform of third-party reviews in China.

In this study we consider the introduction of "Douban Reading" on Tencent.com being an exogenous shock to Douban network. Tencent.com, the largest Facebook-type social network in China, started testing customizable installations of third-party applications on their users' profile pages (called "QZone") on August 1, 2009. The first such application was Douban Reading. Figure 1 is a screenshot of Douban Reading taken from a Tencent user's QZone. This application essentially granted a Tencent user a registered Douban account (or a connection to his or her existing Douban account) upon installation, and the direct access to all Douban book pages (including all book reviews). However, as its name suggests, the application did not provide direct entry into other

[3] Take book pages as an example, Douban provides detailed information such as the ISBN number and a brief introduction. All user reviews with preview sentences are listed underneath the information section. Douban also makes the links to purchase the corresponding products available on product pages. More detailed information can be found in Section 4.

[4] Statistics on the size of Douban.com can be found in the annual surveys prepared by the program "We Are Social" at `http://www.wearesocial.com/`.

product pages, unless the user went to the corresponding Douban pages. Furthermore, this application also made the links to online bookstores, e.g., Amazon.cn, available to Tencent users. Tencent discontinued this application on October 31, 2011. Although the service terminated, Tencent users are able to keep their Douban accounts.

[Insert Figure 1 about here.]

A straightforward impact of the exogenous shock is a sudden explosion in the number of registered users on Douban.com. In Figure 2 we depict the total number and growth rate of registered users on Douban.com. The platform indeed experienced an outburst of its users population. Particularly, the total number of registered users was $3,790,891$ by the end of July 2009 (right before the introduction of Douban Reading), and skyrocketed to $8,466,660$ after a month.[5] The population was more than doubled within a month. The merge between Douban.com and QZone of Tencent.com, through the application Douban Reading, provides us a unique opportunity to evaluate its impacts on individual users' incentives and activities of posting online product reviews.

[Insert Figure 2 about here.]

4. Data and Samples

This section provides detailed description of the data and samples used in our empirical analysis. We start with our sampling process and a brief overview of the original sample. The main sample used in our empirical analysis is introduced in Section 4.2.

4.1. The Original Sample

To evaluate the effect of the exogenous shock on Douban users' contributions, we crawled all relevant information from the webpages of $24,374$ users. The specific sampling process is as follows. We start with a random sample of 119 Douban groups with all their members, originally $552,827$ users. We select $24,374$ users in this sample, because only these users joined Douban.com before the exogenous shock. We crawled all publicly available user demographics on their profile pages,

[5] Douban.com used to publish and update the total number of registered users on its homepage on a daily basis, and removed it after June 2013.We obtain the statistic from the historical data maintained by the Internet Archive, `http://www.archive.org`.

all reviews, and related product information. The reason we focus on this specific subgroup is by design, we study the network size effects on individual users' reviewing behavior. Only this subsample allows us to compare the changes in their onsite activities before and after the exogenous shock.

From the user profile pages, we obtain user demographic information including age, gender, location, and registration date. We dumped all reviews of books, movies, TV shows, and music (if any) for each sampled user. For each product review, we collected the timestamp, title, full content, and all comments attached (with comments posting dates as well). We also obtained the review ratings, ranging from 1 (the worst) to 5 (the best) in integers, the number of helpfulness votes, and the number of non-helpfulness votes for each review.[6] In addition, we crawled all other activities such as the number of photo albums, personal notes, product collections,[7] and public messages. For a user's network information, we obtained the identities of all his or her followers and followees by the data collection date, November 1, 2014.

For each review, we downloaded all product information from the product pages. For instance, if the product is a book, we crawl the following information the author(s), the publisher, ISBN number, publishing date, the total number of pages, a brief outline of the book content, and the author(s) introduction. We also obtained the product rating, ranging from 1 (the worst) to 10 (the best). Note that this rating is assigned by any Douban users that select an integer within the range, not necessarily those who have reviewed the product. Therefore, it is different from the rating assigned in the reviews. As mentioned above, the review rating is chosen by the review writer, with a range 1 to 5.

Table 1 reports summary statistics for this original sample of users and all their reviews by our data collection date. Note that $5,823$ out of $24,374$ users had posted at least one product review

[6] The readers of a product review have the option to characterize the review being either "helpful" or "not helpful." The platform displays the total number for each option.

[7] Douban users can choose to display their own collections of books, movies, TV shows, or music on their profile pages. Products in collections are usually personal favorites including those reviewed by themselves. The collections are divided into three categories. Collections of books, as examples, are categorized into those one has read, those one is reading, and those in one's wish list. Similar categorization applies to other types of products.

(regardless of categories) by November 1, 2014. Among these "active" users, the average number of reviews ever posted is around 7. Among all the reviews, about 36% are in the book section, and more than 50% are movies or TV shows reviews. The rest are music reviews. The average rating had reached over 4 (out of 5) up to November 2014, with the average number of helpful votes greater than 7, and the percentage of helpful votes over 43%. It is noteworthy that the distribution of the volume of reviews is highly skewed, with over half the population contributing nothing. This phenomenon of "specialization" is consistent with theory predictions on the private provision of public goods in social networks (Bramoullé and Kranton 2007).

[Insert Table 1 about here.]

By our data collection date, among all sampled users, the median number of followees and followers are 36 and 33 respectively. Figure 3 shows the distributions of followees count (out-degree) and followers count (in-degree). Both distributions abide by the power law or "fat-tail" phenomenon that is well documented in the literature.[8] Indeed, we notice that the maximum number of followees count is $2,144$ in our sample, and that for followers is more than $88,000$.

[Insert Figure 3 about here.]

4.2. The Sample in Regressions

In our main empirical analysis in Section 5, we focus on the reviews posted by the sampled users between July 4, 2009 and August 28, 2009, which is 4 weeks before and 4 weeks after the exogenous shock. We choose this specific time window mainly because the exogenous shock was the only policy change on Douban.com within this period. Another reason of focusing on a relatively small time window is to control for the "spill over" effect of the exogenous shock. Tencent users who installed the Douban Reading app may gradually start reading non-book reviews, i.e., movie or music reviews. We argue that in our sampling period, a short period of time before and after the merge, this spill-over effect does not significantly affect our comparisons.

[8] Price (1965) was the first to document such distributions in the setting of social networks. Jackson (2010) provides a thorough review and theoretical treatments on this topic.

We first take a closer look at these sampled reviews categorized by types, i.e., either book or other reviews, in Table 2. We notice that more than half of the sampled book reviews were posted during the four weeks after the exogenous shock, while less than 50% of other reviews (including all movie, TV shows, and music reviews) were made in the same period. This suggests that Douban users indeed became more "active" in writing book reviews after the exogenous shock, relative to their activeness in composing other reviews.

[Insert Table 2 about here.]

In the original sample, $23,548$ users registered on Douban.com before July 4, 2009. Notice that there is slight difference from the original sample of $24,374$ users who joined before August 1, 2009. It turns out that these two samples are not significantly different in terms of review activities or demographic information. However, there is a non-negligible fraction of "silent" users in the sample. Specifically, about 16.5% ($3,878$ out of $23,548$) of the sampled users posted at least one review, regardless of the product section, during the study period. The workhorse or the main sample used in the empirical analysis in Section 5 contains these $3,878$ active users. We provide a sample comparison in Table 3. Not surprisingly, these users were indeed more active in posting reviews and more engaged with the platform.[9]

[Insert Table 3 about here.]

5. Empirical Analysis

5.1. Differences-In-Differences Design

Our purpose is to establish the causal effect of a larger group of audience and peer reviewers on Douban users' behavior of posting online reviews. With the exogenous shock, a simple and naïve method is to compare some measures of individual contribution activities, e.g., the weekly number of and the average textual length of reviews posted, before and after the introduction of Douban Reading. A serious concern undermining this naïve method is that the revealed effects confound

[9] This may lead to a concern of sample selection. We conduct a robustness check in Section 5.2.4 by including all $23,548$ users, and the results are qualitatively the same as our main findings. In fact, we are able to reproduce all findings using the larger sample. It is expected since the rest of the users contributed no reviews either before or after the event.

with the overall trend on Douban.com. Specifically, this method hypothetically treats reviews after the exogenous shock as the treatment group, while those prior to the shock the control group. A potential problem is that the difference between the twogroups captures not only the treatment effect (group size effects on contributions), but also the intrinsic difference (such as the overall trend in review posting on Douban.com). Therefore, a simple before-and-after comparison masks the true effect on review contributions.

A quasi-experimental design, the differences-in-differences method (often abbreviated as diffs-in-diffs or DID), provides a solution to the endogeneity problem (Card and Krueger 1994, Athey and Imbens 2006, Chevalier and Mayzlin 2006). The exogenous shock to Douban.com provides us a unique opportunity to apply this method. Suppose Y_i is some measure of user i's contributions to Douban reviews, e.g., the weekly number of reviews. Let Y_i^T denote the user's contributions to Douban book reviews, and Y_i^C his or her contributions to other reviews including those for movies, TV shows, and music.[10] As the superscripts suggest, we divide all Douban reviews into two groups—book reviews as the treated group and other reviews as the untreated/control group.[11] A DID estimator of the treatment effect uses an assumption (usually called the assumption of common trend) that in the absence of treatment the average difference in the outcome variable, Y_i, between the treated and untreated would have stayed roughly constant, in our case, before and after the exogenous shock (Abadie 2005).

Based on the setup and common trend assumption, let $\bar{Y}_s^T$ and $\bar{Y}_s^C$ be the average contributions in period s ($s = 1, 2$) to Douban book reviews (treatment) and other reviews (control), respectively, for our sampled users. Period $s = 1$ indicates the period before the exogenous shock, while period $s = 2$ takes place after the exogenous shock. A simple DID estimator would be to calculate $\hat{\beta} =$

[10] Traditional DID method leaves the choice of control groups to the researchers, prompting critics about the arbitrariness of the selection and the degree to which the control units can credibly proxy for the treated group's counterfactual outcomes. A recent strand of literature (Abadie et al. 2010) propose the synthetic control method to "optimally" select the control units based on pre-treatment characteristics. Our setup can circumvent this issue because we are comparing the same user in the treated group (book reviews) and in the control group (other reviews).

[11] One might have a concern that movie reviews are fundamentally different from music reviews, because the population of movie reviewers might be completely different from those who specialize in music reviews. We conduct a robustness check comparing book reviews to movie reviews and music reviews separately in Section 5.2.4. The findings are not qualitatively different.

$\left(\bar{Y}_2^T - \bar{Y}_2^C\right) - \left(\bar{Y}_1^T - \bar{Y}_1^C\right)$. That is, an *unconditional* version of the DID estimator can be defined as the difference between the difference in average contribution levels between book and other reviews after the exogenous shock, and the same difference for the pre-treated period.

In some instances, the common trend assumption adopted for DID estimator may not be plausible because the treated and untreated differ according to some variables, $\mathbf{X}_{it}$. An example is the user demographics. The rationale is that the treatment effect may differ for different types of users. In this situation, a regression formulation of the DID estimator is useful to compute a *conditional* version that corrects for the effect of $\mathbf{X}_{it}$. Specifically, our main empirical specification is

$$Y_{it}^j = \beta_0 + \beta_1 \cdot D^j \cdot D_t + \beta_2^{'} \mathbf{X}_{it} + \beta_3 \cdot D^j + \mu_i + \nu_t + \epsilon_{it}^j, \tag{1}$$

where the superscript j $(j = T, C)$ indicates corresponding variables for book reviews and other reviews respectively (e.g., Y_{it}^T is a certain measure of user i's contributions to book reviews in period t.); the dummy $D^T = 1$ indicates book reviews; the event dummy $D_t = 1$ $(t = 1, 2, 3, ...)$ stands for the period after the exogenous shock. In addition, $\mathbf{X}_{it}$ include user demographics (both time-variant and time-invariant) such as the length of time on site and user location dummies.

In this equation, β_1 is the coefficient for the interaction term between the dummy for book reviews and the event dummy. It is the coefficient of interests capturing the treatment effect of introducing Douban Reading on some outcome variable, Y_{it}^j. We include user fixed effects, μ_i, and time fixed effects, ν_t, to control for unobserved characteristics at the individual user level and those affecting all users but differing in time. Note that a separate term for D_t is omitted because of multicollinearity with the time dummies ν_t.

5.2. Average Peer Effects

Before presenting the evidence of the group size effects, we first show that the exogenous shock indeed led to a larger population of audience for the sampled reviews. We then start with the evidence showing the distinct behaviors in posting book reviews and other reviews before and after the shock from summary statistics. We continue to report estimation results of the regression-adjusted model (Equation (1) in the last section). We conduct several robustness checks in addition to our main specifications and samples.

5.2.1. Larger Reference Groups A prerequisite of our analysis is whether the introduction of Douban Reading indeed led to a larger population of audience and peer review writers on Douban.com. Unfortunately, we are unable to observe the number of actual review readers. However, a proxy for the population of audience is the quantity of comments made to the reviews. There exist unambiguously positive correlations between the population of audience and the number of comments they post. Thus, controlling for all observed characteristics, we hypothesize that the exogenous shock led to more comments for reviews.

We focus on the comments made to the reviews posted before July 4, 2009. We keep track of all comments between July 4, 2009 and August 28, 2009, which is our main study period. During this period, $1,110$ reviews (out of $20,539$ total reviews) received $2,123$ comments from readers. The average number of comments a review received during the period was 1.39 per week. We conduct similar DID analysis, as in Section 5.1, to explore the effect of the exogenous shock on the number of comments made to an individual review. The dependent variable, Y_{it}^j in Equation (1), is the log of the number of comments per review per week in this set of analysis.

The estimates are reported in Table 4. The estimates of the main effect support our hypothesis. Specifically, the estimate of β_1, reported in the first row of Table 4, suggests that the exogenous shock led to on average about 1% more comments per review per week. The finding implies that the group size of review readers, who posted the comments, was indeed enlarged due to the exogenous shock. This fills the gap between the exogenous shock—the introduction of Douban Reading on Tencent QZone—and the group size effects on review posting behaviors. It is safe to use the event indicator as a proxy for the increase in the size of peer groups.

[Insert Table 4 about here.]

5.2.2. Evidence from Summary Statistics Figure 4 depicts the monthly total number of reviews by categories between August 2008 and August 2010, one year before and one year after the exogenous shock. It clearly shows the difference between the monthly volume of book reviews and other (or movie) reviews around the shock. In fact, the monthly number of book reviews rose from 310 in July 2009 to 322 in August 2009 by our sampled users; in contrast the total number of all other reviews dropped from 589 in July 2009 to 481 in August 2009 (by more than 18%). Similarly, we observe that movie reviews decreased by more than 20% from 454 in July to 355 in August 2009 over the exogenous shock.

[Insert Figure 4 about here.]

To test each of our hypotheses, we construct the dependent variables accordingly as in Table 5. Table 6 presents the detailed summary statistics of all dependent variables by category. For each individual user and each type of reviews, we cluster the weeks before and the weeks after the exogenous shock separately, and compare them using the paired one-sided T-tests. The results, reported in the last two columns, suggest that without any control variables, on average, these users posted more and longer book reviews after August 1, 2009; while their non-book reviews or movie reviews were less and shorter. The average and standard deviation of ratings were both higher after the shock for book reviews only. In addition, the average helpfulness scores were significantly larger in book reviews, but smaller in other categories.

[Insert Table 5 and 6 about here.]

We further compare differences in users' reviewing behaviors before and after. As an example, we first calculate, for each user in a week, the difference between the total number of book reviews and that of non-book (or movie) reviews before the exogenous shock. Then we compute the same difference after the exogenous shock. Eventually, we compare these two sets of difference using paired one-sided T-tests. We replicate similar comparisons for other dependent variables. The first two comparisons, as shown in the panel of "Difference in # reviews" and "Difference in total lengths of reviews" in Table 6, suggest that compared to the difference prior to the shock, the difference afterwards was significantly bigger (in negative values) implying that the individuals contributed more to book reviews relative to other reviews after the exogenous shock. Similarly, we find that both the difference in the average ratings and the difference in the standard deviation of ratings were bigger significantly after the shock, illustrated in the panel "Difference in the avg. review rating" and "Difference in the s.d. of the review ratings" of Table 6. Similarly, the difference in the average percentage of helpfulness votes (%) was also significantly bigger, as shown in the last panel of Table 6.

5.2.3. Evidence from Regression-Adjusted Analysis Table 7 reports the estimation results from Equation (1).[12] Each column corresponds to a hypothesis in Section 2.

[Insert Table 7 about here.]

We first report the estimates from regressions using the weekly number of reviews as a measure of contribution levels. The results are presented in the first column of Table 7. From the semi-log specifications, the coefficient estimate of the key interaction term (indicating book reviews after the exogenous shock) suggests that, conditional on all control variables, compared to the difference between the number of book reviews and that of other reviews before the exogenous shock, the same difference afterwards rose by about 0.6% (supporting Hypothesis H1a). As an alternative evaluation of individual users' contributions to product reviews, the number of characters in a review also reflects the effort the contributor devotes into the provision of the public good. We thus consider an alternative specification using the log of textual lengths as the dependent variable. Estimation results are presented in the second column of Table 7. The coefficient estimate shows that compared to the difference between the weekly total lengths of book reviews and those of other reviews before the shock, the same difference after the exogenous shock increased by around 4.2% (supporting Hypothesis H1b).

In addition to the volume effects, we also estimate the causal effects on the valence of reviews. Column 3 of Table 7 presents the estimation results with the dependent variable the average ratings assigned by an individual reviewer in a week. Consistent with Hypothesis H2, the estimate suggests that the sampled Douban users assigned, on average, higher ratings for book reviews than other reviews conditional on all observed characteristics. The precise marginal effect is around 0.03, and

[12] For all estimations, there might be concerns of serial correlations since the observations are weekly records in a continuous period. In all estimations we report the robust standard errors as suggested by Wooldridge (2002). We obtain similar results by calculating either bootstrapped standard errors or standard errors clustering at the user level.

statistically significant. It can be interpreted that the exogenous shock caused individuals to assign, on average, 0.03 point of ratings higher (with the range between 1 and 5).

We report the estimation results for the test of Hypothesis H3 in the fourth column of Table 7. The findings support the hypothesis that the variance, or the standard deviation, of weekly review ratings assigned by an individual user is indeed higher for the book section than in other sections. Similarly, the calculated causal effect of the exogenous shock is about 0.001. In other words, the exogenous shock increased the average variation of review ratings by about 0.001 in standard deviation, which is statistically significant.

The last set of estimation results are presented in the last column of Table 7, used to test the hypotheses H4a and H4b. The positive estimate of β_1 lends support to H4a, suggesting that a representative Douban user tends to post more helpful reviews in a larger network (or in front of a larger population of audience and cohort reviewers). Quantitatively, the exogenous shock led to about 0.13% increase in the percentage of helpful votes.

The findings above support all our hypotheses. Various concerns may mitigate the reliability of these findings, including concerns about our samples and empirical specifications. In the following section, we conduct robustness checks (or placebo tests) according to each of those concerns.

5.2.4. Robustness Checks and Placebo Tests In our main specification, we stack all other reviews together and treat them as the control group. In fact, this group is composed of reviews of movies (including TV shows) and music. The behavior of posting reviews might be different in different segments. As seen in our summary statistics, more than half of Douban reviews are posted in the movie section. One might have a concern that combining all other reviews together may mask the true difference, between individual behaviors devoted to book reviews and other reviews. In the first set of specification tests, we investigate whether our main findings would be attenuated suppose we compare book reviews to other reviews separately. Specifically, we perform two separate comparisons, i.e., book reviews versus movie reviews and book reviews versus music reviews. Estimation results are reported in Table 8 and 9. We find qualitatively consistent estimates with our main findings in Table 7.

[Insert Table 8 and 9 about here.]

The second concern about our main specification is that one may worry about the sampling period. In our main estimations, we focus on a period between July 4, 2009 and August 28, 2009. One may worry about the validity of our results from this particular period. In response, we conduct robustness checks based on alternative samples that cover different periods of time. Specifically, we construct two additional samples that include reviews between June 6, 2009 and September 25, 2009 (8 weeks before and 8 weeks after the exogenous shock), and between May 9, 2009 and November 22, 2009 (12 weeks before and after the shock) respectively. We report estimation results in Table 10 and 11. We find the estimates qualitatively consistent with our main findings again. We also notice that the magnitudes of the coefficient estimates for the contribution levels, the log of reviews quantity and textual lengths, are decreasing in the length of time covered in the study periods, i.e., the effect is more significant with a tighter period around the exogenous shock.

[Insert Table 10 and 11 about here.]

Third, the main sample comprises of users that registered before July 4, 2009. One may worry that individuals joining the platform earlier are qualitatively different from those joining late. For instance, the early users, who were still active in our study period, tend to be more loyal to the platform than those that enter late. So their behavior of writing reviews might be significantly different from the newcomers. To make sure that our findings are consistent across different groups of users in terms of their maturity with the platform, we conduct a third set of robustness checks. In particular, we divide our original sample to two subsample. The first subsample contains all users that joined before July 4, 2008, which is one year before July 4, 2009; the second subsample includes all other users. We repeat our main estimations using these two subsamples separately, and report the results in Table 12 and 13. We find consistent results as our main findings again. We also notice that the exogenous shock has greater impacts on the individuals that entered later than those joining earlier.

[Insert Table 12 and 13 about here.]

We conduct a placebo test to make sure that it is indeed the exogenous shock that is driving our results, not by coincidence. Specifically, if the exogenous shock had no effects at all on individual users' behavior, we would expect similar results—positive effects on volume/valence/helpfulness—by treating a random date, other than August 1, 2009, as the date of the exogenous shock. We look at two such dates, one before the exogenous shock and the other afterwards. We choose July 4, 2009 and August 29, 2009 as the two dates. For each date, we repeat our estimations by focusing on the same length of time as in our main specification, four weeks before and four weeks after the "fake" event date. For example, for July 4, 2009, we compare the changes from June 6, 2009 to July 31, 2009. We report the results in Table 14 and 15. Clearly, both "fake" exogenous shock dates had no effects at all on review volume/valence/helpfulness. These findings provide strong evidence that it is indeed the exogenous shock causing the behavioral changes.

[Insert Table 14 and 15 about here.]

As mentioned in Section 4, there exists a non-negligible fraction of users contributing no reviews during our study period. There are even "zombie" users who had never posted any reviews by our data collection date in 2014. As a last robustness check, we ask whether including these inactive users can change our results qualitatively. We report the same set of estimations in Table 16. Although all estimates are much smaller in absolute values, the directions and significance levels are highly comparable to our main results as in Table 7. This is not unexpected since by our empirical strategy, the DID design, we are essentially comparing a user's activities before and after the shock. There was no difference at all in inactive users' activities before and after. Therefore, the estimated causal effects are smaller in magnitude, but the directions should follow.

[Insert Table 16 about here.]

5.3. The Moderating Roles of User Characteristics

So far we have identified the average peer effects in the sense that we have not explored the heterogeneity across different types of users. As an example, it is not hard to see that the sudden expansion in the users population can have differential effects on the users who are experts in writing book reviews, versus those interested in reviewing movies only, since the shock led to an increase in the audience of book reviews mainly. For a comprehensive understanding of the impacts of the exogenous shock, we continue exploring the moderating role of user characteristics. Specifically, we study three important aspects about an individual review writer, namely the activeness or the engagement with the platform, their expertise, and popularity. The estimation equation is an adaption from Equation (1).

$$Y_{it}^j = \beta_0 + \beta_1 \cdot D^j \cdot D_t \cdot X_i^{\text{mod}} + \beta_2 \cdot D^j \cdot D_t + \beta_3^{'} \mathbf{X}_{it} + \beta_4 \cdot D^j + \mu_i + \nu_t + \epsilon_{it}^j, \tag{2}$$

where in addition to all variables in Equation (1), X_i^{mod} is a certain characteristic about the user i. In the new estimation equation, β_1 has a different interpretation. The estimates now capture the differential effects of the exogenous shock by the user characteristic X_i^{mod}. Note again that a single term of X_i^{mod} is omitted because of its collinearity with the user fixed effects μ_i.

We start with the activeness of a user, or his or her engagement with the platform. A more engaged user typically spends more time on the platform. He or she may contribute more reviews (regardless of the product section), and engages more in other onsite activities as well. It is to be expected that these users will be less affected by a sudden change in the population of audience or cohort review writers, because their contributions are less casual compared to less active users. In this regard, we hypothesize that the exogenous shock has smaller effects on a user's contributions in either the volume, the valence, or the quality.

We adopt two measures of a user's activeness (or engagement with the platform): the number of product reviews (regardless of the product type) prior to our study period and the number of followees. Our estimation results are reported in Table 17 and 18 respectively. The estimates of β_1 in various regressions, shown in the first rows of the two tables, support our conjecture of less impacts on more active users. In particular, a user, who had posted more reviews before the beginning of our study period, has a lower level of the increase in contribution volume, the average and dispersion of ratings, as well as the rise in the quality of reviews. We find similar patterns for the other moderating characteristic, the number of followees.

[Insert Table 17 and 18 about here.]

Second, Douban users can have distinct interests in reviewing different types of products. A "bookworm" may have interests in writing reviews for books only, while a movie addict is only knowledgable for and can contribute to movie reviews. The application Douban Reading, facilitating direct access to the book section on Douban.com, apparently increased the population of audience and review writers for books mainly. In contrast, the users who focused more on reviews in other categories did not experience the same level of growth in their reference groups. Due to these reasons, we expect to see larger impacts of the merge event on users specialized in book reviews, while in contrast limited impacts on those who are experts in other categories.

To formally test this conjecture, we construct two user characteristics measuring their expertise, namely the number of book reviews and other reviews prior to the study period separately. We estimate Equation (2) with these two moderating characteristics separately, and the results are reported in Table 19 and 20. As expected, the estimates of β_1 (of Equation (2)) for the number of book reviews are mostly positive and statistically significant, suggesting that the exogenous shock had bigger impacts on the "bookworms." While in sharp contrast, the moderating effects of the number of reviews in other categories are uniformly negative at any usual significance level.

[Insert Table 19 and 20 about here.]

In the last set of tests of the moderating effects, we look at the role of a user's popularity. With the similar reasoning as the moderating role of one's activeness, we hypothesize that the introduction of Douban Reading had smaller impacts on a more popular user. We use the number of followers as the measure of popularity, and repeat the estimations of Equation (2). Results are reported in Table 21. We find that the moderating effects of the followers count are indeed negative, but not statistically significant except the effect on the quality of reviews.

[Insert Table 21 about here.]

6. Concluding Remarks

The management of online consumer responses is becoming increasingly important for companies to utilize social media (Gu and Ye 2014, Sun and Xu 2016). As an emerging form of consumer responses (word of mouth), online reviews, provided by either consumers or third-party professionals, are closely related to customer satisfaction and product sales. We study individuals' incentives to post online reviews in the first place, and focus on the fundamental factor governing their contributions—the peer effects. We complement the recent literature on the antecedents of online reviews by establishing the causal effects of peer groups (Goes et al. 2014, Wang et al. 2015, Khern-am-nuai and Kannan 2016).

We examine a large-scale online networking platform, Douban.com, which allows its users to publish product reviews. The identification of the causal effect of peer groups hinges on a sudden and unexpected merge with a large-scale and influential online social network—an exogenous shock to the Douban network. Using this "natural experiment," we conduct DID analysis and find that, first of all, the exogenous shock caused Douban users to write more and longer reviews. Not limited to the effects on the volume of contributions, we also find statistically significant effects on the ratings (valence of reviews). Specifically, the exogenous shock caused individual users to assign higher and more extensive ratings. Also consistent with our hypothesis, we find that users posted reviews of higher quality (more helpful for review readers) following the merge event. Even more interestingly, we document the moderating roles of user characteristics. In particular, a user, who was more active on (or more engaged with) the platform, more popular, and less knowledgable, was less affected by the exogenous shock. In other words, these characteristics all have negative moderating effects.

Our results have important implications for companies as well as online platforms that rely on users' contributions. Understanding consumers' incentives to post online reviews and how their behaviors are affected by online friends/audience/followers lies at the core of any company's management of online customer responses. Our empirical findings suggest that firms can improve the ratings of their product or service, and encourage consumers to post more helpful responses (as purchasing guidance for potential customers) by taking strategies to enlarge the population of the review audience and peer review writers. Furthermore, our findings of the moderating effects recommend that companies should focus on targeted groups of consumers that post online reviews.

For example, companies should encourage less active review writers to speak up, and pay more attention to the reviewers that are more knowledgable in their own products or service. For online platforms that rely on user contributions, since their ultimate goal is to encourage users to speak up, they should target at a larger users base, and pay particular attention to "small" users that are less active or less popular.

Our study can be extended in several ways. First, in the current study, we focus on the network size effects. Future studies can explore more from the perspective of network positions. It will be interesting to study the dynamics between a user's online contributions and his or her positions in a network. Along this line of research, future studies can explore the interrelationships between a user's contributions and his or her neighbors' (online friends, followers, or audience) contributions. Second, we notice that less than 25% of the sampled users ($5,823$ out of $24,374$) have ever posted any reviews in our data. As a future direction, one can explore more on the skewness of contributions in online communities, and how to mitigate this issue to encourage more active contributions from the whole community. Last but not least, beyond the specific context in the current study, future research can apply similar identification strategies, the DID analysis of similar "natural experiments," to address the endogeneity issue in social network studies.

Figures and Tables:

Figure 1 **An Illustration of "Douban Reading" Installed on a Tencent User's QZone**

Figure 2 **The Total Number and Growth Rate of Registered Users on Douban.com**

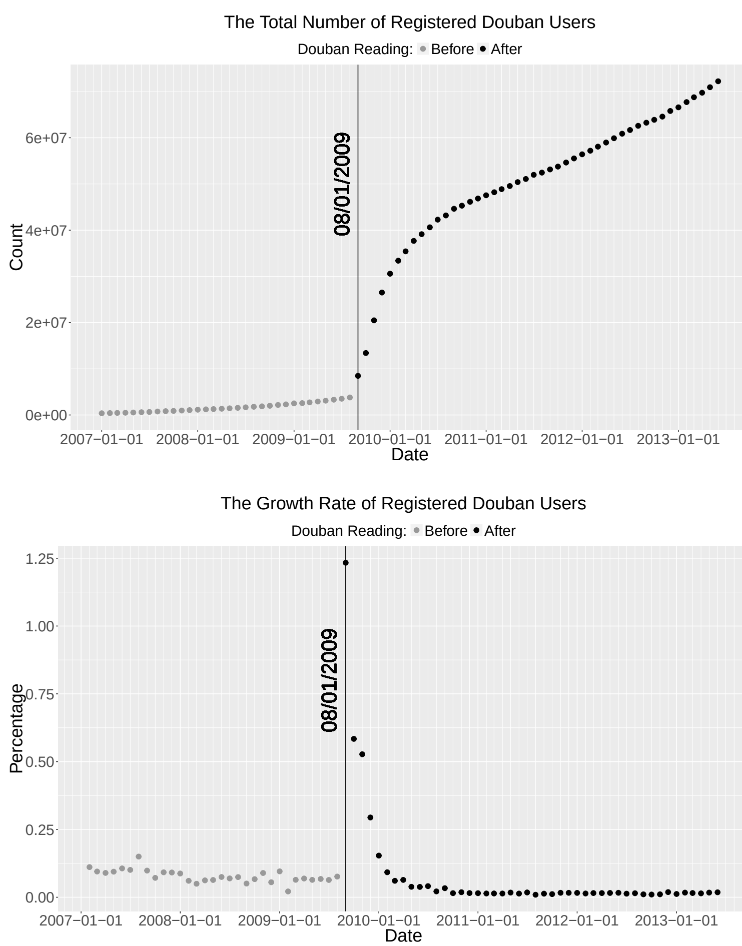


Figure 3 Distributions of the Followees and Followers Count for Users in the Main Sample

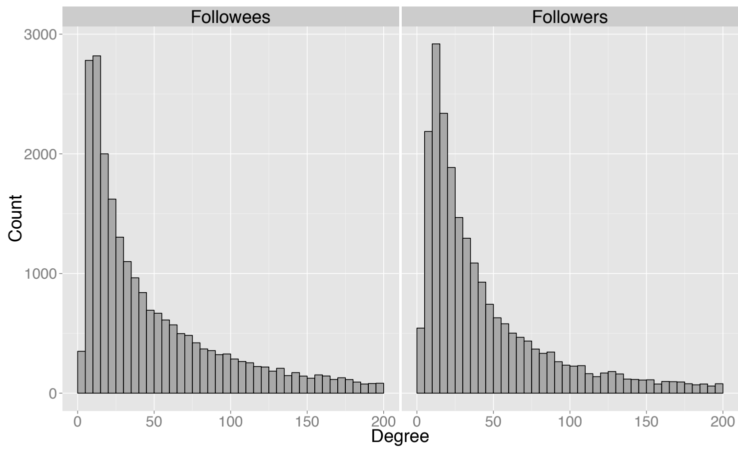
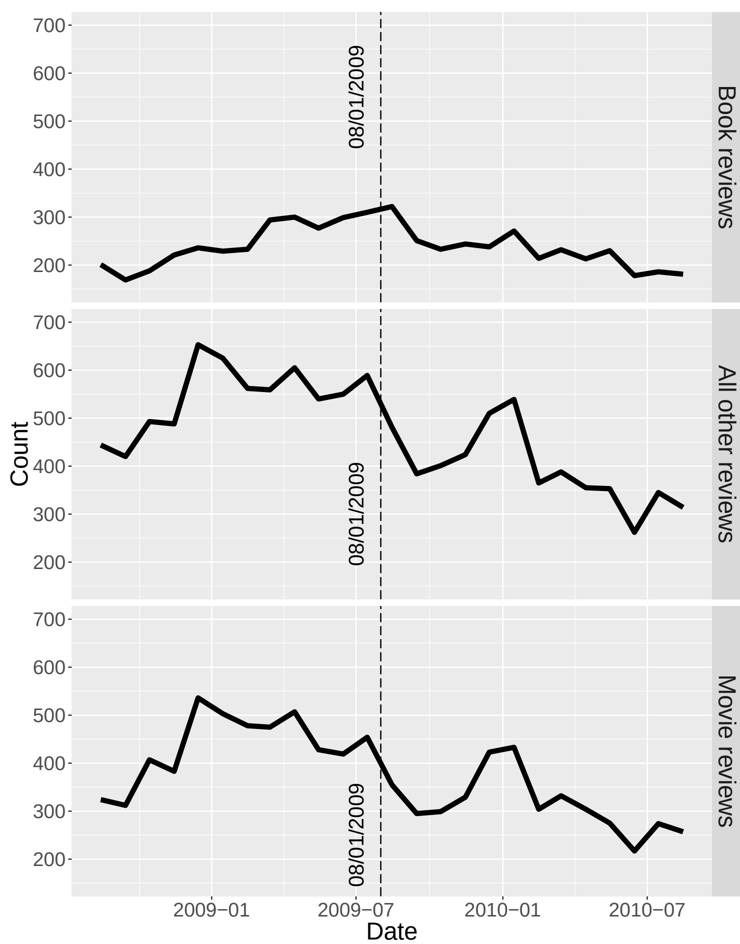


Figure 4 Monthly Total Numbers of Reviews by Review Categories



Summary Statistics:

Table 1 Summary Statistics of All Sampled Users and Reviews

	Statistics				
Variables	Median	Mean	s.d.	Min.	Max.
User Information					
# All reviews	0	2.570	12.603	0	881
# Groups	43	66.536	68.274	0	2,004
# Followees	36	78.926	142.174	0	2,144
# Followers	33	158.171	1,517.220	0	88,656
1 (Missing followers)	0	0.002	0.044	0	1
Length on site[a]	424	498.221	365.539	1	1,613
Book Collections					
# Reading	2	6.556	36.526	0	3,217
# Wish list	10	53.530	230.823	0	15,908
# Read	21	65.255	142.938	0	5,236
Movie Collections					
# Watching	0	3.164	8.101	0	245
# Wish list	13	72.420	204.302	0	9,530
# Watched	99	247.785	377.361	0	7,201
Music Collections					
# Listening	1	7.587	35.956	0	2,784
# Wish list	2	18.053	133.431	0	16,287
# Listened	10	87.472	323.902	0	10,049
Users			24,374		
Review Information					
Length of review[b]	513	859.241	1,966.795	0	297,620
# Helpful votes	1	7.739	76.175	0	10,324
# Non-helpful votes	0	1.009	6.162	0	455
Helpfulness votes (%)[c]	20	43.836	45.845	0	100
Rating (1 - 5)	4	4.050	0.939	1	5
1 (Missing rating)	0	0.018	0.135	0	1
1 (After Douban Reading)	0	0.489	0.500	0	1
Users			5,823		
Book reviews			14,969		
Movie reviews			21,150		
Music reviews			5,575		
All reviews			41,774		

[a] The number of days, since a user's registration, till Aug. 1, 2009, the date of the exogenous shock.

[b] The length of the review content, not including title.

[c] The score equals #helpful votes / (#helpful votes + #non-helpful votes).

Table 2 Summary Statistics of the Sampled Reviews by Product Type

	Summary Statistics					
	Book reviews		Other reviews			
			All other reviews[a]		Movie reviews	
Variables:	Mean	s.d.	Mean	s.d.	Mean	s.d.
Length of review	884.490	1,483.970	845.141	2,190.490	824.766	2,239.410
# Helpful votes[b]	6.023	30.385	8.698	92.331	9.530	103.241
# Non-helpful votes	0.747	3.606	1.156	7.202	1.261	7.982
Helpfulness votes (%)[c]	48.753	46.264	41.090	45.378	38.438	44.789
Rating (1 - 5)	3.929	1.132	4.000	1.046	3.934	1.057
1 (Missing rating)	0.030	0.169	0.012	0.110	0.011	0.103
1 (After Douban Reading)	0.554	0.497	0.451	0.498	0.47	0.499
Num. obs.	14,969		26,806		21,150	

[a] All reviews other than those in book section, including movie, TV shows, and music reviews.

[b] The number of "helpful" clicks by review readers, likewise for "# Non-helpful votes."

[c] The score equals #helpful votes / (#helpful votes + #non-helpful votes).

Table 3 Sample Comparisons

	Sample in regressions			Original sample		
Variables:	Median	Mean	s.d.	Median	Mean	s.d.
# All reviews	4	9.416	26.023	0	2.570	12.603
# Groups	52	75.865	71.113	43	66.536	68.274
# Followees	51	100.862	161.432	36	78.926	142.174
# Followers	52	229.683	1,840.202	33	158.171	1,517.22
1 (Missing followers)	0	0.003	0.056	0	0.002	0.044
Length on site[a]	648	688.208	382.911	424	498.221	365.539
Users		3,878			24,374	

[a] The number of days, since a user's registration, till Aug. 1, 2009, the date of the exogenous shock.

Table 4 Exogenous Shock Effects on the Number of Comments

Dep var.: log (# Comments)	OLS estimates
$1\,(\mathbf{Tencent}) * 1\,(\mathbf{Book})$	$\mathbf{0.001}^{**}$
	(0.408e-03)[a]
Control variables[b]	Yes
Review fixed effects	Yes
Week fixed effects	Yes
1 (Book)	Yes
Adj. R^2	0.020
Num. obs.	328,624

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

[a] We report the robust standard errors in all estimation results. One may argue that it is likely to overestimate the precision of estimates with the robust standard errors. Therefore in all estimations, we also check the results with standard errors clustered at individual user level. The significance results are not surprisingly similar.

[b] We control for the length of time the review has been posted, and the squared term of the length.

Table 5 A Summary of Dependent Variables

Hypothesis	Dependent variable (weekly)
H1A	Log of the number of reviews
H1B	Log of the total textual lengths of reviews
H2	Average review ratings
H3	Standard deviation of review ratings
H4	Percent of helpfulness votes (%)[a]

[a] The measure is constructed from the votes by review readers, not assigned by the review writers. Specifically, all readers can vote a review being "helpful" or "not helpful." This variable is equal to the percentage of the "helpful" votes among all votes.

Empirical Evidence:

Table 6 Comparing Individuals' Review Activities Before and After – Summary Statistics

	Before Aug. 1, 2009			After Aug. 1, 2009				
Variables: (weekly[a])	Median	Mean	s.d.	Median	Mean	s.d.	t-stats[b]	p-values
Book reviews								
Avg. # reviews	0	0.017	0.147	0	0.019	0.109	-0.438	0.331
Avg. textual length	0	14.924	190.950	0	17.109	279.944	-0.402	0.344
Avg. rating	0	0.048	0.265	0	0.060	0.299	-1.949	0.026
Avg. s.d. of rating	0	0.001	0.019	0	0.002	0.026	-0.655	0.256
Avg. helpfulness votes (%)[c]	0	0.643	4.775	0	0.800	5.156	-1.395	0.082
Non-book reviews								
Avg. # reviews	0	0.034	0.184	0	0.027	0.168	1.758	0.961
Avg. textual length	0	24.609	213.531	0	20.516	214.079	0.843	0.800
Avg. rating	0	0.099	0.386	0	0.082	0.346	2.092	0.982
Avg. s.d. of rating	0	0.003	0.033	0	0.002	0.026	1.293	0.902
Avg. helpfulness votes (%)	0	0.967	5.435	0	0.858	4.961	0.921	0.822
Difference in avg. # reviews								
Book – non-book[d]	0	-0.016	0.208	0	-0.008	0.180	-1.880	0.030
Difference in avg. lengths of reviews								
Book – non-book	0	-9.685	283.251	0	-3.407	349.026	-0.870	0.192
Difference in avg. review rating								
Book – non-book	0	-0.051	0.427	0	-0.021	0.414	-3.132	<0.001
Difference in the avg. s.d. of review rating								
Book – non-book	0	-0.001	0.037	0	-0.0002	0.036	-1.478	0.070
Difference in avg. helpfulness votes (%)								
Book – non-book	0	-0.324	6.683	0	-0.058	6.703	-1.752	0.040
Users				3,878				
Weeks				8				
Num. obs.				62,048				

[a] We evaluate all variables on a weekly basis. Our sampling period is between July 4 and August 28, 2009, spanning an 8-week window before and after the shock.

[b] In the paired one-sided T-tests our null hypotheses are that the mean value of the variable before the exogenous shock is greater than that after the shock.

[c] The measure is constructed from the votes by review readers, not assigned by the review writers. Specifically, all readers can vote a review being "helpful" or "not helpful." This variable is equal to the percentage of the "helpful" votes among all votes.

[d] This variable calculates, for each user, the difference between the average value of the variable (across four weeks) before the exogenous shock and that after the shock.

Table 7 Main Results – Estimates of the Exogenous Shock Effects

	OLS estimates				
Dep var.:	log (#Reviews)	log (Rev. length)	Avg. rating	SD ratings	Helpful votes (%)
$1(\mathbf{Tencent}) * 1(\mathbf{Book})$	**0.006*****	**0.042*****	**0.030*****	**0.001***	**0.266****
	(0.002)[a]	(0.013)	(0.008)	(0.001)	(0.133)
Control variables[b]	Yes	Yes	Yes	Yes	Yes
User fixed effects	Yes	Yes	Yes	Yes	Yes
Week fixed effects	Yes	Yes	Yes	Yes	Yes
1 (Book)	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.138	0.109	0.091	0.052	0.079
Num. obs.	62,048	62,048	62,048	62,048	62,048

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

[a] We report the robust standard errors in all estimation results. One may argue that it is likely to overestimate the precision of estimates with the robust standard errors. Therefore in all estimations, we also check the results with standard errors clustered at individual user level. The significance results are not surprisingly similar.

[b] We control for a user's experience with Douban platform by including the length of time on site, and the squared term of the length. This applies to all subsequent estimations.

Robustness Checks and Placebo Tests:

Table 8 Robustness – Comparing Book Reviews with Movie Reviews

	OLS estimates				
Dep var.:	log (#Reviews)	log (Rev. length)	Avg. rating	SD ratings	Helpful votes (%)
1 (**Tencent**) ∗ 1 (**Book**)	**0.005*****	**0.039*****	**0.029*****	**0.001***	**0.290****
	(0.002)[a]	(0.012)	(0.005)	(0.001)	(0.123)
Control variables	Yes	Yes	Yes	Yes	Yes
User fixed effects	Yes	Yes	Yes	Yes	Yes
Week fixed effects	Yes	Yes	Yes	Yes	Yes
1 (Book)	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.128	0.103	0.085	0.056	0.076
Num. obs.	62,048	185,760	62,048	62,048	62,048

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

[a] We report the robust standard errors in all estimation results. One may argue that it is likely to overestimate the precision of estimates with the robust standard errors. Therefore in all estimations, we also check the results with standard errors clustered at individual user level. The significance results are not surprisingly similar.

Table 9 Robustness – Comparing Book Reviews with Music Reviews

	OLS estimates				
Dep var.:	log (#Reviews)	log (Rev. length)	Avg. rating	SD ratings	Helpful votes (%)
1 (**Tencent**) ∗ 1 (**Book**)	**0.003****	**0.021****	**0.015****	**0.001**	**0.138**
	(0.001)[a]	(0.009)	(0.006)	(0.001)	(0.103)
Control variables	Yes	Yes	Yes	Yes	Yes
User fixed effects	Yes	Yes	Yes	Yes	Yes
Week fixed effects	Yes	Yes	Yes	Yes	Yes
1 (Book)	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.114	0.089	0.075	0.042	0.067
Num. obs.	62,048	185,760	62,048	62,048	62,048

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

[a] We report the robust standard errors in all estimation results. One may argue that it is likely to overestimate the precision of estimates with the robust standard errors. Therefore in all estimations, we also check the results with standard errors clustered at individual user level. The significance results are not surprisingly similar.

Table 10 Robustness – An Alternative Sample: Jun 6, 2009 – Sep. 25, 2009

	OLS estimates				
Dep var.:	log (#Reviews)	log (Rev. length)	Avg. rating	SD ratings	Helpful votes (%)
$1(\mathbf{Tencent}) * 1(\mathbf{Book})$	**0.004*****	**0.024*****	**0.019*****	**0.001***	**0.132**
	(0.001)[a]	(0.009)	(0.006)	(0.001)	(0.091)
Control variables	Yes	Yes	Yes	Yes	Yes
User fixed effects	Yes	Yes	Yes	Yes	Yes
Week fixed effects	Yes	Yes	Yes	Yes	Yes
1 (Book)	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.112	0.090	0.074	0.038	0.064
Num. obs.	123,968	123,968	123,968	123,968	123,968

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

[a] We report the robust standard errors in all estimation results. One may argue that it is likely to overestimate the precision of estimates with the robust standard errors. Therefore in all estimations, we also check the results with standard errors clustered at individual user level. The significance results are not surprisingly similar.

Table 11 Robustness – An Alternative Sample: May 9, 2009 – Oct. 23, 2009

	OLS estimates				
Dep var.:	log (#Reviews)	log (Rev. length)	Avg. rating	SD ratings	Helpful votes (%)
$1(\mathbf{Tencent}) * 1(\mathbf{Book})$	**0.003*****	**0.019*****	**0.019*****	**0.001****	**0.072**
	(0.001)[a]	(0.007)	(0.005)	(0.416e-03)	(0.072)
Control variables	Yes	Yes	Yes	Yes	Yes
User fixed effects	Yes	Yes	Yes	Yes	Yes
Week fixed effects	Yes	Yes	Yes	Yes	Yes
1 (Book)	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.092	0.075	0.062	0.024	0.055
Num. obs.	185,760	185,760	185,760	185,760	185,760

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

[a] We report the robust standard errors in all estimation results. One may argue that it is likely to overestimate the precision of estimates with the robust standard errors. Therefore in all estimations, we also check the results with standard errors clustered at individual user level. The significance results are not surprisingly similar.

Table 12 Robustness – Users Joining before July 4, 2008

	OLS estimates				
Dep var.:	log (#Reviews)	log (Rev. length)	Avg. rating	SD ratings	Helpful votes (%)
1 (**Tencent**) ∗ 1 (**Book**)	**0.003***	**0.026***	**0.017***	**0.001**	**0.271***
	(0.002)[a]	(0.014)	(0.009)	(0.001)	(0.145)
Control variables	Yes	Yes	Yes	Yes	Yes
User fixed effects	Yes	Yes	Yes	Yes	Yes
Week fixed effects	Yes	Yes	Yes	Yes	Yes
1 (Book)	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.123	0.094	0.081	0.061	0.073
Num. obs.	46,080	46,080	46,080	46,080	46,080

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

[a] We report the robust standard errors in all estimation results. One may argue that it is likely to overestimate the precision of estimates with the robust standard errors. Therefore in all estimations, we also check the results with standard errors clustered at individual user level. The significance results are not surprisingly similar.

Table 13 Robustness – Users Joining after July 4, 2008

	OLS estimates				
Dep var.:	log (#Reviews)	log (Rev. length)	Avg. rating	SD ratings	Helpful votes (%)
1 (**Tencent**) ∗ 1 (**Book**)	**0.012*****	**0.088*****	**0.067*****	**0.002**	**0.253**
	(0.004)[a]	(0.030)	(0.020)	(0.002)	(0.301)
Control variables	Yes	Yes	Yes	Yes	Yes
User fixed effects	Yes	Yes	Yes	Yes	Yes
Week fixed effects	Yes	Yes	Yes	Yes	Yes
1 (Book)	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.158	0.132	0.103	0.043	0.090
Num. obs.	15,968	15,968	15,968	15,968	15,968

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

[a] We report the robust standard errors in all estimation results. One may argue that it is likely to overestimate the precision of estimates with the robust standard errors. Therefore in all estimations, we also check the results with standard errors clustered at individual user level. The significance results are not surprisingly similar.

Table 14 Estimates of the Exogenous Shock Effects – A "Fake" Shock on July 4, 2009

	OLS estimates				
Dep var.:	log (#Reviews)	log (Rev. length)	Avg. rating	SD ratings	Helpful votes (%)
1 (**Tencent**) ∗ 1 (**Book**)	**-0.002**	**-0.019**	**-0.011**	**-0.001**	**-0.137**
	(0.002)[a]	(0.013)	(0.008)	(0.001)	(0.131)
Control variables	Yes	Yes	Yes	Yes	Yes
User fixed effects	Yes	Yes	Yes	Yes	Yes
Week fixed effects	Yes	Yes	Yes	Yes	Yes
1 (Book)	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.114	0.098	0.080	0.046	0.069
Num. obs.	61,984	61,984	61,984	61,984	61,984

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

[a] We report the robust standard errors in all estimation results. One may argue that it is likely to overestimate the precision of estimates with the robust standard errors. Therefore in all estimations, we also check the results with standard errors clustered at individual user level. The significance results are not surprisingly similar.

Table 15 Estimates of the Exogenous Shock Effects – A "Fake" Shock on August 29, 2009

	OLS estimates				
Dep var.:	log (#Reviews)	log (Rev. length)	Avg. rating	SD ratings	Helpful votes (%)
1 (**Tencent**) ∗ 1 (**Book**)	**-0.001**	**-0.010**	**-0.006**	**0.000**	**-0.188**
	(0.002)[a]	(0.012)	(0.008)	(0.001)	(0.126)
Control variables	Yes	Yes	Yes	Yes	Yes
User fixed effects	Yes	Yes	Yes	Yes	Yes
Week fixed effects	Yes	Yes	Yes	Yes	Yes
1 (Book)	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.127	0.098	0.082	0.037	0.070
Num. obs.	61,984	61,984	61,984	61,984	61,984

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

[a] We report the robust standard errors in all estimation results. One may argue that it is likely to overestimate the precision of estimates with the robust standard errors. Therefore in all estimations, we also check the results with standard errors clustered at individual user level. The significance results are not surprisingly similar.

Table 16 Robustness – With "Inactive" Users

	OLS estimates				
Dep var.:	log (#Reviews)	log (Rev. length)	Avg. rating	SD ratings	Helpful votes (%)
$1\,(\mathbf{Tencent}) * 1\,(\mathbf{Book})$	**0.941e-03*****	**0.688e-04*****	**0.493e-04*****	**0.201e-03***	**0.044****
	(0.298e-03)[a]	(0.002)	(0.001)	(0.121e-03)	(0.022)
Control variables	Yes	Yes	Yes	Yes	Yes
User fixed effects	Yes	Yes	Yes	Yes	Yes
Week fixed effects	Yes	Yes	Yes	Yes	Yes
1 (Book)	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.149	0.120	0.102	0.053	0.085
Num. obs.	376,768	376,768	376,768	376,768	376,768

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

[a] We report the robust standard errors in all estimation results. One may argue that it is likely to overestimate the precision of estimates with the robust standard errors. Therefore in all estimations, we also check the results with standard errors clustered at individual user level. The significance results are not surprisingly similar.

Moderating Effects:

Table 17 Moderating Effects by the Volume of Early Reviews

	OLS estimates				
Dep var.:	log (#Reviews)	log (Rev. length)	Avg. rating	SD ratings	Helpful votes (%)
1 (**Tencent**) ∗ 1 (**Book**) ∗#**Early reviews**	**-0.001***** (0.289e-03)[a]	**-0.005***** (0.002)	**-0.003***** (0.001)	**-0.196e-03**** (0.795e-04)	**-0.050***** (0.016)
1 (Tencent) ∗ 1 (Book)	0.011*** (0.002)	0.070*** (0.014)	0.045*** (0.009)	0.002*** (0.001)	0.529*** (0.150)
Control variables	Yes	Yes	Yes	Yes	Yes
User fixed effects	Yes	Yes	Yes	Yes	Yes
Week fixed effects	Yes	Yes	Yes	Yes	Yes
1 (Book)	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.141	0.110	0.092	0.052	0.080
Num. obs.	62,048	62,048	62,048	62,048	62,048

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

[a] We report the robust standard errors in all estimation results. One may argue that it is likely to overestimate the precision of estimates with the robust standard errors. Therefore in all estimations, we also check the results with standard errors clustered at individual user level. The significance results are not surprisingly similar.

Table 18 Moderating Effects by the Number of Followees

	OLS estimates				
Dep var.:	log (#Reviews)	log (Rev. length)	Avg. rating	SD ratings	Helpful votes (%)
1 (**Tencent**) ∗ 1 (**Book**) ∗#**Followees**	**-0.158e-04**** (0.793e-07)[a]	**-0.103e-03*** (0.607e-04)	**-0.451e-04** (0.415e-04)	**-0.194e-07** (0.286e-07)	**-0.001** (0.001)
1 (Tencent) ∗ 1 (Book)	0.007*** (0.002)	0.052*** (0.014)	0.034*** (0.009)	0.001* (0.001)	0.381*** (0.145)
Control variables	Yes	Yes	Yes	Yes	Yes
User fixed effects	Yes	Yes	Yes	Yes	Yes
Week fixed effects	Yes	Yes	Yes	Yes	Yes
1 (Book)	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.138	0.109	0.091	0.052	0.079
Num. obs.	62,048	62,048	62,048	62,048	62,048

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

[a] We report the robust standard errors in all estimation results. One may argue that it is likely to overestimate the precision of estimates with the robust standard errors. Therefore in all estimations, we also check the results with standard errors clustered at individual user level. The significance results are not surprisingly similar.

Table 19 Moderating Effects by the Number of Early Book Reviews

	OLS estimates				
Dep var.:	log (#Reviews)	log (Rev. length)	Avg. rating	SD ratings	Helpful votes (%)
1 (**Tencent**) $*$ 1 (**Book**) $*$#**Early book reviews**	**0.001**** (0.001)[a]	**0.009*** (0.004)	**0.005*** (0.002)	**0.343e-03** (0.260e-03)	**0.094*** (0.050)
1 (Tencent) $*$ 1 (Book)	0.004** (0.001)	0.026* (0.014)	0.022** (0.009)	0.001 (0.001)	0.111 (0.143)
Control variables	Yes	Yes	Yes	Yes	Yes
User fixed effects	Yes	Yes	Yes	Yes	Yes
Week fixed effects	Yes	Yes	Yes	Yes	Yes
1 (Book)	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.139	0.109	0.091	0.052	0.079
Num. obs.	62,048	62,048	62,048	62,048	62,048

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

[a] We report the robust standard errors in all estimation results. One may argue that it is likely to overestimate the precision of estimates with the robust standard errors. Therefore in all estimations, we also check the results with standard errors clustered at individual user level. The significance results are not surprisingly similar.

Table 20 Moderating Effects by the Number of All Other Reviews Previously

	OLS estimates				
Dep var.:	log (#Reviews)	log (Rev. length)	Avg. rating	SD ratings	Helpful votes (%)
1 (**Tencent**) $*$ 1 (**Book**) $*$#**Early other reviews**	**-0.002***** (0.315e-03)[a]	**-0.009***** (0.002)	**-0.005***** (0.001)	**-0.312***** (0.086e-03)	**-0.080***** (0.017)
1 (Tencent) $*$ 1 (Book)	0.011*** (0.002)	0.073*** (0.014)	0.046*** (0.009)	0.002*** (0.001)	0.558*** (0.142)
Control variables	Yes	Yes	Yes	Yes	Yes
User fixed effects	Yes	Yes	Yes	Yes	Yes
Week fixed effects	Yes	Yes	Yes	Yes	Yes
1 (Book)	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.144	0.111	0.092	0.053	0.081
Num. obs.	62,048	62,048	62,048	62,048	62,048

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

[a] We report the robust standard errors in all estimation results. One may argue that it is likely to overestimate the precision of estimates with the robust standard errors. Therefore in all estimations, we also check the results with standard errors clustered at individual user level. The significance results are not surprisingly similar.

Table 21 Moderating Effects by the Number of Followers

	OLS estimates				
Dep var.:	log (#Reviews)	log (Rev. length)	Avg. rating	SD ratings	Helpful votes (%)
1 (**Tencent**) $*$ 1 (**Book**) $*$#**Followers**	**-0.226e-07** (0.151e-07)[a]	**-0.166e-04** (0.128e-04)	**-0.667e-07** (0.885e-07)	**-0.708e-09** (0.489e-09)	**-0.266e-03*** (0.137e-03)
1 (Tencent) $*$ 1 (Book)	0.006*** (0.002)	0.046*** (0.013)	0.031*** (0.009)	0.001* (0.001)	0.327** (0.134)
Control variables	Yes	Yes	Yes	Yes	Yes
User fixed effects	Yes	Yes	Yes	Yes	Yes
Week fixed effects	Yes	Yes	Yes	Yes	Yes
1 (Book)	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.138	0.109	0.091	0.052	0.079
Num. obs.	62,048	62,048	62,048	62,048	62,048

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

[a] We report the robust standard errors in all estimation results. One may argue that it is likely to overestimate the precision of estimates with the robust standard errors. Therefore in all estimations, we also check the results with standard errors clustered at individual user level. The significance results are not surprisingly similar.

References

Abadie, A. 2005. Semiparametric Difference-in-Differences Estimators. *Review of Economic Studies* **72**(1): 1–19.

Abadie, A., A. Diamond, and J. Hainmueller. 2010. Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. *Journal of the American Statistical Association* **105**(490): 493–505.

Abrahams, A. S., W. Fan, G. A. Wang, Z. Zhang, and J. Jiao. 2015. An Integrated Text Analytic Framework for Product Defect Discovery. *Production and Operations Management* **24**(6): 975–990.

Andreoni, J. 1989. Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence. *Journal of Political Economy* **97**(6): 1447–1458.

Andreoni, J. 2007. Giving Gifts to Groups: How Altruism Depends on the Number of Recipients. *Journal of Public Economics* **91**(9): 1731–1749.

Andreoni, J., and B. D. Bernheim. 2009. Social Image and the 50-50 Norm: A Theoretical and Experimental Analysis of Audience Effects. *Econometrica* **77**(5): 1607–1636.

Athey, S., and G. Imbens. 2006. Identification and Inference in Nonlinear Difference-In-Differences Models. *Econometrica* **74**(2): 431–497.

Bateman, P. J., P. H. Gary, and B. S. Butler. 2011. The Impact of Community Commitment on Participation in Online Communities. *Information Systems Research* **22**(4): 841–854.

Bramoullé, Y. and R. Kranton. 2007. Public Goods in Networks. *Journal of Economic Theory* **135**(1): 478–494.

Cao, Q., W. Duan, and Q. Gan. 2011. Exploring Determinants of Voting for the "Helpfulness" of Online User Reviews: A Text Mining Approach. *Decision Support Systems* **50**(2): 511–521.

Card, D., and A. B. Krueger. (1994). Minimum Wages and Employment: A Case Study of the Fast Food Industry in New Jersey and Pennsylvania. *American Economic Review* **84**(4): 772–793.

Chamberlin, J. 1974. Provision of Collective Goods as a Function of Group Size. *American Political Science Review* **68**(2): 707–716.

Chen, H., Z. Zheng, and Y. Ceran. (2016). De-Biasing the Reporting Bias in Social Media Analytics. *Production and Operations Management* **25**(5): 849–865.

Chen, Y., and J. Xie. (2005). Third-Party Product Review and Firm Marketing Strategy. *Marketing Science* **24**(2): 218–240.

Chevalier, J. A., and D. Mayzlin. (2006). The Effect of Word of Mouth on Sales: Online Book Reviews. *Journal of Marketing Research* **43**(3): 345–354.

Dellarocas, C. 2003. The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms. *Management Science* **49**(10): 1407–1424.

Duan, W., B. Gu, and A. B. Whinston. 2008. Do Online Reviews Matter? An Empirical Investigation of Panel Data. *Decision Support Systems* **45**(4): 1007–1016.

Duncan, B. 2004. A Theory of Impact Philanthropy. *Journal of Public Economics* **88**(9-10): 2159–2180.

Forman, C., A. Ghose, and B. Wiesenfeld. 2008. Examining the Relationship Between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets. *Information Systems Research* **19**(3): 291–313.

Goes, P., M. Lin, and C. Au Yeung. 2014. "Popularity Effect" in User-Generated Content: Evidence from Online Product Reviews. *Information Systems Research* **25**(2): 222–238.

Goldsmith-Pinkham, P., and G. W. Imbens. 2013. Social Networks and the Identification of Peer Effects. *Journal of Business and Economic Statistics* **31**(3): 253–264.

Gu, B., P. Konana, R. Rajagopalan, and H-WM Chen. 2007. Competition among Virtual Communities and User Valuation: The Case of Investing-Related Communities. *Information Systems Research* **18**(1): 68–85.

Gu, B., and Q. Ye. 2014. First Step in Social Media: Measuring the Influence of Online Management Responses on Customer Satisfaction. *Production and Operations Management* **23**(4): 570–582.

Harsanyi, J. C. 1980. Rule Utilitarianism, Rights, Obligations and the Theory of Rational Behavior. *Theory and Decision* **12**(2): 115–133.

Jackson, M. O. 2010. Social and Economic Networks. *Princeton University Press*.

Khern-am-nuai, W., and K. Kannan. 2016. How do Monetary Incentives Affect Online Reviews? *Working Paper.*

Lacetera, N. and M. Macis. 2010. Social Image Concerns and Prosocial Behavior: Field Evidence from a Nonlinear Incentive Scheme. *Journal of Economic Behavior and Organization* **76**(2): 225–237.

Lee, Y.-J., K. Hosanagar, and Y. Tan. 2015. Do I Follow My Friends or the Crowd? Information Cascades in Online Movie Ratings. *Management Science* **61**(9): 2241–2258.

Li, X., and L. M. Hitt. 2008. Self-Selection and Information Role of Online Product Reviews. *Information Systems Research* **19**(4): 456–474.

Liu, Y. 2006. Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue. *Journal of Marketing* **70**(3): 74–89.

Manski, C. F.. 1993. Identification of Endogenous Social Effects: The Reflection Problem. *Review of Economic Studies* **60**(3): 531–542.

Mudambi, S. and D. Schuff. 2010. What Makes a Helpful Review? A Study of Customer Reviews on Amazon.com. *MIS Quarterly* **34**(1): 185–200.

Ofir, C., and I. Simonson. 2001. In Search of Negative Customer Feedback: The Effect of Expecting to Evaluate on Satisfaction Evaluations. *Journal of Marketing Research* **38**(2): 170–182.

Olson, M. 1965. The Logic of Collective Action: Public Goods and the Theory of Groups. Harvard University Press.

Peng, J., A. Agarwal, K. Hosanagar, and R. Iyengar. 2016. Network Embeddedness and Content Sharing on Social Media Platforms. *Working Paper.*

Price, D. J. S. 1965. Networks of Scientific Papers. *Science* **149**: 510–515.

Ransbotham, S., G. C. Kane, and N. H. Lurie. 2012. Network Characteristics and the Value of Collaborative User-Generated Content. *Marketing Science* **31**(3): 387–405.

Sun, H., and L. Xu. 2016. Online Reviews and Collaborative Service Provision: A Signal-Jamming Model *Production and Operations Management.* Forthcoming.

Sun, M. 2012. How Does the Variance of Product Ratings Matter? *Management Science* **58**(4): 696–707.

Wang, C., X. Zhang, and I. Hann. 2015. Socially Nudged: A Quasi-Experimental Study of Friends' Social Influence in Online Product Ratings. *Information Systems Research*. Forthcoming.

Wei, Z., and M. Xiao. 2016. For Whom to Tweet? Evidence from a Large-Scale Social Media Platform. *Working Paper*.

Wooldridge, J. 2002. Econometric Analysis of Cross Section and Panel Data. The MIT Press.

Zhang, X., and F. Zhu. 2011. Group Size and Incentives to Contribute: A Natural Experiment at Chinese Wikipedia. *American Economic Review* **101**(4): 1601–1615.

Zhu, F., and X. Zhang. 2010. Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics. *Journal of Marketing* **74**(2): 133–148.