



Information Systems Research

Publication details, including instructions for authors and subscription information:
<http://pubsonline.informs.org>

“Popularity Effect” in User-Generated Content: Evidence from Online Product Reviews

Paulo B. Goes, Mingfeng Lin, Ching-man Au Yeung

To cite this article:

Paulo B. Goes, Mingfeng Lin, Ching-man Au Yeung (2014) “Popularity Effect” in User-Generated Content: Evidence from Online Product Reviews. Information Systems Research 25(2):222-238. <https://doi.org/10.1287/isre.2013.0512>

Full terms and conditions of use: <https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2014, INFORMS

Please scroll down for article—it is on subsequent pages



With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

“Popularity Effect” in User-Generated Content: Evidence from Online Product Reviews

Paulo B. Goes, Mingfeng Lin

Eller College of Management, University of Arizona, Tucson, Arizona 85721
{pgoes@eller.arizona.edu, mingfeng@eller.arizona.edu}

Ching-man Au Yeung

Axon Labs Ltd., Hong Kong, albertaueung@gmail.com

Online product reviews are increasingly important for consumer decisions, yet we still know little about how reviews are generated in the first place. In an effort to gather more reviews, many websites encourage user interactions such as allowing one user to subscribe to another. Do these interactions actually facilitate the generation of product reviews? More importantly, what kind of reviews do such interactions induce? We study these questions using data from one of the largest product review websites where users can subscribe to one another. By applying both panel data and a flexible matching method, we find that as users become more popular, they produce more reviews and more *objective* reviews; however, their numeric ratings also systematically change and become more negative and more varied. Such trade-off has not been previously documented and has important implications for both product review and other user-generated content websites.

Keywords: product reviews; user-generated content; online community; opinion leader; social media; popularity; text mining; matching

History: Ram Gopal, Senior Editor; Ming Fan, Associate Editor. This paper was received on September 5, 2012, and was with the authors 4 months for 1 revision. Published online in *Articles in Advance* January 30, 2014.

1. Introduction

Online user reviews have become an increasingly important source of information for consumers. Many studies have shown that online reviews significantly affect consumer choices and product sales (Berger et al. 2010, Chintagunta et al. 2010, Dellarocas et al. 2007, Duan et al. 2008, Forman et al. 2008, Li and Hitt 2008, Liu 2006, Sohl 1999, Sorenson and Stuart 2005, Sun 2012, Zhu and Zhang 2010). However, much less is known about how reviews are generated in the first place. Several prior studies emphasized consumers' independent decisions to write reviews. For instance, they are more likely to post reviews online when they are very satisfied or very dissatisfied, resulting in bimodal distributions of user ratings (Hu et al. 2006). Some recent studies emphasize how ratings may affect each other, in the sense that expressed opinions of others may influence future opinions (Moe and Trusov 2011).

We study how users' online interactions may also affect the generation of user ratings. Although ratings have been shown to influence each other (Moe and Trusov 2011), most online users are silent (Dellarocas and Wood 2008) and only considered passive readers of reviews. As many websites become increasingly “social,” those who read reviews and those who write reviews can easily connect with, and therefore

influence, each other. Readers are no longer just a fuzzy collective that receives others' opinions but are also becoming individually visible to those who write. For instance, readers may “like” a review or share it on social media sites, they may rate it as helpful or unhelpful, or they can subscribe to select writers' future writings. In other words, sites that incorporate social media features allow unprecedented ease for writers to keep track of their audience or “fans.” For this reason, even if readers are silent, they can still influence the behavior of writers. This is consistent with the well-known Hawthorn Effect (Adair 1984), where the mere presence of observers can change behaviors. For user reviews, however, there have been no published studies documenting the presence or absence of such an effect, yet it directly determines the credibility and usefulness of these reviews, especially because of the fast growth of online social media. We therefore ask the following research question:

How does the interaction among online users influence their review-writing behavior, including the frequency of writing, the opinions that they express, and how they express them?

We focus on one popular type of user online interaction, i.e., subscription or “following.” When a reader subscribes to a writer, content generated by

that writer will have priority over other writers when displayed to the reader. For review writers, therefore, their subscribers (followers) constitute essentially a captive audience, and our goal is to understand how such an audience affects the behavior of the review writers. Given the widely recognized importance of user reviews for consumer decision making, and the prevalence of user interactions on product review websites, it is critical that we have a better understanding of whether such effects exist. If user interactions encourage them to write certain types of reviews but discourage others, then this trade-off should be carefully weighed, and any induced "bias" should be recognized. Websites may have to balance the need to generate more reviews and the need to avoid potential biases. For firms that are trying to evaluate their products' market response, or consumers trying to make purchase decisions, such effects should be taken into account as well.

In the next section, we provide an overview of our empirical context, *epinions.com*. Section 3 reviews the literature related to our study and develops specific hypotheses. In §4 we discuss the data that we use for the empirical analysis, as well as our empirical strategy. Section 5 presents a discussion of the results from our analyses. In §6, we discuss the implications of this study and some directions for future research.

2. Context

We obtain data from *epinions.com* to empirically study how opinion writers' behaviors change as their audience grows. *Epinions.com* is uniquely ideal for the purpose of our study because of the availability of details on product reviews, the presence of directional subscription ties between users,¹ and the time stamp for each tie. These features allow us to construct a longitudinal data set that includes objective measures of consumer interactions (especially the number of incoming ties from peers) and product reviews, so we can examine how users change their product review behavior as they gather a virtual following. In this section, we briefly describe how *epinions.com* works, especially as relevant to our research question.²

Epinions.com is one of the largest websites dedicated to product reviews on the Internet. It allows users to search for products, read reviews and ratings from other consumers, and optionally contribute their own reviews. The reviews on this site include product reviews (containing textual opinions and a numeric product rating) as well as generic articles that are not

targeted at a specific product. We refer to the latter as "nonrating articles" in this paper.³ What makes the website particularly interesting is the "web-of-trust" (WOT) feature. Each user can choose to "trust" one or more other users, so content written by that trusted user will be given higher priority when displayed. For instance, if a user John reads Jane's reviews and like them, John can trust Jane by clicking a link on her profile. This tie does not require approval from Jane, and Jane does not have to reciprocate it (i.e., trusting John back) either. Once such a tie is created, if John searches for information about a product that Jane had written a review for, Jane's review for that product will be displayed to John ahead of other reviews. This is highly comparable to following a user on Twitter or other social media sites. In addition, John may trust other users, and Jane may trust other users as well; this in turn creates a directed WOT network of site users. The goal of our paper is to study how user interactions on this network, especially the number of incoming ties that a user receives, affect users' behavior in writing product reviews. In the next section, we review some related theoretical and empirical studies and derive a set of testable hypotheses.

3. Related Literature and Hypotheses

3.1. Related Literature

We draw on several important and growing streams of research in the literature: *social influence*, *online word-of-mouth*, and *online communities*. Our objective is not to be exhaustive in including all papers written in these areas but to highlight those that directly inform our analyses and to discuss the gap in the existing literature that we seek to fill.

Existing empirical studies of social influence in information systems have largely focused on behavioral similarities, i.e., the behavior of one person influencing another that they are connected to. For example, many researchers study how peer behavior influences the adoption of products and services (Aral et al. 2009, Iyengar et al. 2011). In the online social media context, Susarla et al. (2012) show that social influence affects how popular YouTube videos can become. An arguably special case of social influence is "opinion leaders," where some members of the population may exert a disproportionately high level of influence on others' product choices. Various methods have been proposed to identify opinion leaders in a network (Iyengar et al. 2011, Trusov et al. 2010). Using data from *epinions.com*, Lu et al. (2013) provide insights into how ties are formed over time and how opinion leaders emerge.

¹ Such ties are referred to as "trust" on *epinions.com*, but they are highly similar to the subscription ties on other social media sites.

² These features are accurate as of the time that data were collected (July 2009).

³ As an example, a nonrating article could be a generic purchase guide (e.g., for buying a car).

Our study can potentially fill a remaining gap in this burgeoning literature. Although the concept of opinion leaders implies that these leaders have their own independent opinions, it may not be the case on social media sites where the generators and consumers of content interact with each other. Social psychologists argue that the mere presence of observers can change behaviors, in what is referred to as the Hawthorne Effect (Adair 1984). Similarly in online social media, subscriptions from users allow a content generator to keep track of the size of their audience. More important, these ties indicate a degree of trust in the writers because they allow the writers to easily "push" their writings to the followers. Given the presence and trust of such an audience, therefore, it seems natural that the review writers' behaviors may be affected. For product reviews, incoming ties may affect the writer's decision on whether to write, how much to write, what to write, and how to write it.

The other stream of literature that we draw on is online word-of-mouth (WOM). Many studies have examined how online word-of-mouth, especially in the form of online product reviews and ratings, influences a wide range of outcomes such as consumer choices, product sales, and even investor decisions (Aggarwal et al. 2012, Dellarocas 2003, Duan et al. 2008, Forman et al. 2008, Liu 2006, Sohl 1999, Sorenson and Stuart 2005, Sun 2012, Zhu and Zhang 2010). As we mentioned earlier, we still know little about the generation of product ratings in the first place. Existing research in this area can be largely classified into two categories. One examines the generation of product reviews as an individual consumer decision or a reflection of consumer characteristics. For instance, consumers are more likely to post reviews when they are very happy or very unhappy with a product, which results in the bimodal distribution of online ratings (Hu et al. 2006). Earlier consumers of a product tend to be more zealous about it, so over time, average ratings tend to decrease (Hu et al. 2006, Li and Hitt 2008). Cheema and Kaikati (2010) show that consumers' needs for "uniqueness" may also affect their decision to provide reviews. Using an experimental approach, Rice (2012) finds that the level of uncertainty surrounding transactions can have an influence on rating behaviors as well. Many studies also find that product characteristics such as price (Li and Hitt 2008), popularity (Zhu and Zhang 2010), and market positioning (such as niche versus hit products; Dellarocas et al. 2010) all influence the generation of reviews. A common theme in this literature is that the decision to contribute reviews is a result of consumer characteristics or product characteristics that led to varying consumer experience.

A second category of studies focuses on how expressed product opinions may affect other opinions—for instance, how earlier ratings influence later ratings. Moe et al. (2011) show this effect using data from a retailer's online sales. Chen et al. (2010) show that in an online community, once users know the median number of reviews contributed by other community members, those who used to write less than the median will write more. More recently, Wang et al. (2010) show that among users who are friends, the ratings that they write influence each other as well. Our study takes a new perspective on the generation of product reviews. We study the effect of an online audience on a user's product review behavior, where the subscribers are mostly strangers rather than friends and silent consumers of reviews rather than producers. More important, the social ties connecting the writer and the following are directional, instead of reciprocated friendship ties. Even though the followers do not produce reviews themselves, their presence and actions (trusting the writer) may still influence the behavior of opinion writers in terms of *how much* they produce (volume of reviews) and *what* they produce (valence, variance, and text features of reviews). If such effects exist, it will indicate a new driver of online content generation that has not been identified to date in the literature.

More broadly, we also draw on a growing literature on online communities (Butler and Wang 2012, Faraj and Johnson 2011, Gu et al. 2007, Ransbotham and Kane 2011). Product review sites can be considered a special case of online communities, and many studies in the online communities literature focus on the incentive of users to voluntarily contribute efforts where there is no monetary return, which directly informs our hypothesis regarding the frequency of writing product reviews (Bateman et al. 2011, Ma and Agarwal 2007, Wasko and Faraj 2005). In particular, Wasko and Faraj (2005) find that reputation-seeking motivations and structural embeddedness are two important motivations for users to contribute. Both of them exist only in a social context, as one member relates to another of the same community. Hence, even though the empirical context in those studies may differ from *epinions.com*,⁴ the motivation to seek recognition from others is still likely to play a role. On *epinions.com*, the size of followership is an important indicator for reputation and embeddedness and should therefore affect the behavior of the review writers. On the other hand, our study is

⁴ Empirical studies of online communities often focus on discussion boards, online forums, and online wikis, where users closely interact with each other through asking, answering, and participating in "social productions." By contrast, the interaction on *epinions.com* through web-of-trust is much more subtle since interactions such as trusting someone occur after a review is written.

unique in several ways compared to existing studies in this literature. For instance, many users in our context are silent and passive consumers of content, but we find that their relations to content generators (even when not reciprocated) can still indirectly influence the content being generated. In addition, whereas many studies of online communities use the ask-and-response between users to construct social networks, social relations in our context are much more objective. We further link user interactions to linguistic features of content being produced. In many ways, therefore, we contribute to the growing literature in online communities.

To sum up, our study is related to but strikingly different from existing studies on social influence (or peer effects) as well as existing studies on the antecedents to online word-of-mouth. To our knowledge, this is one of the first to examine how user interactions, particularly those that involve the silent followers on user-generated review sites, may affect the behavior of how users express their opinions online.

3.2. Hypotheses Development

We now develop the main hypotheses that we will test in this paper. Whereas existing studies examine word-of-mouth from the perspective of products, we examine it from the perspective of review *writers* who generate them. Specifically, we study how user interactions affect the (1) volume, or the number of reviews; (2) valence, or the mean of ratings; (3) variance of ratings; and (4) textual features of reviews that users generate online. All these dimensions are important characteristics of online product reviews and have been shown to influence consumer decisions in different ways.

The first metric of interest is the volume of reviews. It is natural to expect that users with a larger online audience should be more likely to contribute more reviews. Research has shown that the act of sharing one's experience with others is largely a public good because of its positive externality (Bolton et al. 2004, Chen et al. 2010): the cost to write is solely borne by the writer, yet readers can read the work without paying the writer. On the other hand, each incoming subscription tie on *epinions.com* suggests that a peer member finds the writer's article to be worthy of reading and trusts that the writer will continue to provide useful information in the future. Given the same degree of externality, the reputation and recognition help internalize some incentive to write. Therefore, receiving ties should encourage the writer to contribute more product reviews. This is also consistent with findings of prior research related to the effect of group size on public goods contribution (Zhang and Zhu 2011). Although it is possible that there may be certain "complacency" effects when the writer has

already reached an "expert" status, it is unlikely to bear first-order consequences on behaviors: only very few will be in that status, so this effect is unlikely to apply to the majority of users.⁵ On the other hand, the encouragement effect of incoming ties should be stronger at the beginning when there are only few followers; an additional 10 subscribers should matter more for someone with only 8 subscribers than for someone with 800 subscribers already. We therefore hypothesize the following:

HYPOTHESIS 1 (H1). Receiving more incoming ties should increase the number of product reviews and non-review articles that a user contributes to the community. However, the marginal effect of more incoming ties should be decreasing.

The valence of ratings that users produce may also be influenced by the presence of an audience, and the effect is likely to be negative. The first reason is related to Hypothesis 1. If users are more likely to evaluate products when they become more popular, that alone may induce a negativity bias (Gu et al. 2007). Through lab experiments and field studies, Ofir and coauthors (2001) show that when consumers are expected to provide an evaluation, they are more likely to focus on negative aspects of a product, resulting in more negative reviews. Hence, all else equal, if H1 is supported, the valence of reviews is likely to decrease in the process. A second reason lies in the behavioral bias of readers toward negative reviews. Previous studies suggest that readers tend to view negative opinions as being more useful or smart (Amabile 1983, Bateman et al. 2011, Moe and Trusov 2011). Yet typical users on the site may not be aware of such behavioral bias from the very beginning. As they gather a larger following, however, they are more likely to recognize such a bias, so they will become more likely to post negative reviews. All else equal, the effect of increasing the number of followers on a review writer's ratings should be negative: the more followers they have, the more likely that the writer will provide negative reviews. Meanwhile, as the number of incoming ties increases, the need to "act smart" will decrease, and the writer may have increasingly lower incentives to post negative reviews. Hence, there should be nonlinearity in the effect as well. We hence hypothesize the following:

HYPOTHESIS 2 (H2). An increasing number of followers will reduce the overall valence of ratings provided by the review writer. The marginal effect, however, should be decreasing.

⁵ In our robustness tests, we remove those with extremely high numbers of incoming ties, to ensure that these outliers or influential observations do not unduly bias our results.

Variance (measured as standard deviations) of ratings, on the other hand, has only recently been recognized as containing valuable information about products (Sun 2012). If the long-term trend of the valence of ratings is decreasing, as we hypothesize above, then given the nature and range of possible ratings on epinions.com, it is likely that the variance of ratings will increase, but at a decreasing rate. A key reason is that epinions.com uses a star-rating system, the highest rating being five stars and the lowest being one star. Hence, if the average rating within a time period is high (close to five stars), then the variance is likely to be small because of the upper limit of five stars. If the average rating within a time period is low (close to one star), then the variance is also likely to be small because of the lower bound of one star. Another reason is that as users become established on the site, they are more likely to review a wider range of products, resulting in a larger variation than before. But with the five-star ratings system, it is not possible for such variation to monotonically increase. Hence, as an extension of the previous hypothesis and the inherent feature of the star ratings system, we hypothesize the following:

HYPOTHESIS 3 (H3). *The variance of reviews that users generate will be higher with more incoming subscription ties, but at a decreasing rate.*

In addition to these characterizations of the numeric ratings, we are also interested in how incoming ties affect the linguistic features of reviews since readers of product reviews look not only at the numerical ratings but also at the textual content of the reviews. To quantify product reviews, a natural metric that we are interested in is their readability, or how easy it is for readers to understand them. As we hypothesized earlier, when a product review writer gathers a larger audience, he or she is likely to write more reviews. An increasing frequency of writing serves as more opportunities to observe feedback and continually practice. By observing readers' responses, review writers can further refine how they write; all else equal therefore, they should be more likely to write better. Hence, we should expect that the overall readability of texts should increase as a user becomes more popular.

A second dimension of product review texts that we are interested in is the degree of objectivity (opposite of emotionality), and we hypothesize that as users become more popular, they are likely to use fewer emotional words in their writings. Two different streams of literature inform this hypothesis. The first one is the Functional Role Theory of sociology (Biddle 1986). Users who receive many incoming ties are *de facto* expert figures in the WOT community. As experts in a product review community, their

role is to provide useful and objective information for readers rather than indulge in emotional rhetoric. According to the functional role theory, these users will recognize such expectations and act accordingly. All else equal, objective writing is more likely to carry a sense of expertise than emotional rants, so we expect these writers to use less emotional words as they become more popular. The second is the social networks literature in sociology and management. To a large degree, the number of incoming ties determines the network position of a user on the WOT. Sociology and management studies suggest that through a social influence process within a network (Marsden and Friedkin 1993), users on similar network positions will have "similar role demands, and similar expectations from others (Ibarra and Andrews 1993)" (Shah 2000, p. 103). In other words, such expectations and demands from other members can motivate them to write in a manner consistent with their "expert" status. We therefore hypothesize the following:

HYPOTHESIS 4 (H4). *As a user's popularity increases, his or her writings' readability should increase, and the use of emotional words should decrease.*

We next turn to the data that we use to test the above hypotheses.

4. Data

We created automated agents to collect data on epinions.com. The first step of our data collection was to identify users on its WOT network. To this end, we employed a snowball approach. Specifically, we started with the top 10 contributors in each main category of epinions.com. For each of these members, we identified all users that trust them and all users that they each trust. These first-degree neighbors were added to the list of users. We then went to these first-degree neighbors' profile pages and found all the members that they trust and members that trust them (second-degree neighbors). We repeated this process so that unique new members IDs were continuously added to the list, until the list no longer grew.⁶ We also obtained the date on which each tie was created;⁷ information about users on their profile pages; and information about the product reviews that they wrote (e.g., time stamp, length of reviews, numeric ratings, etc.). At the end, the data set contains 92,094 user names (all users connected to the WOT),

⁶ Users who are not connected to any members of the WOT, therefore, will not be sampled. However, given the nature and purpose of the site, those users are likely to be few.

⁷ Consistent with prior studies of epinions.com (Lu et al. 2013), we focus on ties created after 2002 because the dates on WOT ties formed prior to 2002 are not available. We obtained the data in July 2009. Some features of the site have since changed.

Table 1 Summary Statistics of User Information (Cross-Sectional, per User)

Variable	Mean	Minimum	Maximum	Standard deviation	<i>N</i>
(a) Number of ratings	11.07	0	4,094	51.36	92,094
(b) Average rating given	3.98	1	5	0.91	62,344
(c) Standard deviation of ratings given	0.79	0	2.83	0.66	62,344
(d) Number of nonrating articles written	1.26	0	438	7.61	92,094
(e) In-degree (trusted by)	6.60	0	2,829	36.31	92,094
(f) Out-degree (trusting)	6.60	0	1,830	27.56	92,094
(g) Number of days since registration	3,041.91	0	3,667	699.49	92,094

Note. This table reports some statistics of the major variables at the end of our data collection time (July 9th, 2009), summarized across users.

Table 2 Correlation (Cross-Sectional)

	Number of ratings	Average rating	Standard deviation of ratings	Number of nonrating articles	In-degree	Out-degree	Days on site
Number of ratings	1.000						
Average rating	−0.001	1.000					
Standard deviation of ratings	0.111	−0.225	1.000				
Number of nonrating articles	0.553	0.013	0.093	1.000			
In-degree	0.638	0.006	0.086	0.557	1.000		
Out-degree	0.458	0.023	0.103	0.448	0.615	1.000	
Days on site	0.020	0.092	0.067	0.070	0.071	0.061	1.000

Note. This table reports cross-correlation among the major variables.

608,047 directional ties on the WOT network, and information on 958,232 reviews. Table 1 provides variable definitions and summary statistics, and Table 2 provides correlation among them.

We refer to the number of incoming ties as the user's in-degree, and the number of outgoing ties as the user's out-degree. Since these ties are directional, there may be overlaps. We refer to the number of nonreciprocated incoming ties as the user's "pure in-degree" and the number of nonreciprocated outgoing ties as the user's "pure out-degree."

For our empirical tests we employ panel data as well as matching methods to test the robustness of our findings. In particular, the matching method not only allows us to estimate the treatment effect of having incoming ties but also accounts for different levels of treatment *intensity* nonparametrically.

5. Models and Results

5.1. Rationale for Modeling Strategies

We are interested in how users change behaviors when they receive incoming user subscription ties, i.e., when they become more popular in the WOT. Hence, our main dependent variables include (1) the number of ratings provided by the user, (2) the mean (or valence) of these ratings, (3) the standard deviation of these ratings, and (4) the number of nonrating articles written by the user. Our main independent variable of interest, is a user's pure in-degree, the number of nonreciprocated incoming ties. Although

the proportion of reciprocated ties is small, those ties may be inherently different from the nonreciprocated ties. The number of reciprocated ties is included in the models as a control variable. Nevertheless, our results are robust even if we do not make this distinction. The number of nonreciprocated outgoing ties, as well as the quadratic term of the above three measures, are all included as control variables. We also control for the length of time since registration to eliminate time effects and include individual fixed-effects.

A cross-sectional analysis of the data could yield erroneous findings because it does not account for potential endogeneity. For instance, although we are interested in understanding how incoming ties affect user behavior in writing reviews, it may be because of the characteristics of their reviews (number, valence, and volume) that earned the trust of other users and resulted in those incoming ties (see Lu et al. 2013). We therefore turn to panel data models and complement them with matching methods.

Some unique features of our context and data help assure the validity of our empirical approaches. First, reverse causality and simultaneity bias are mitigated in the panel data because we measure the network metrics (number of incoming ties) *prior* to the measurement of behaviors such as the number and valence of ratings (Singh et al. 2011). Second, the timing of *incoming* ties is largely exogenous in our context because unlike blog posts, product reviews are shown to users only when they search for the reviewed product, so the timing of a potential reader seeing a review

and then subscribing to its writer is quasi-random. No matter what writers write, there is no guarantee that other users will start following them immediately after they write it. It is also for this reason that we make the distinction between nonreciprocated incoming ties and reciprocated ties because unobservable information is more likely to be present with the latter.

5.2. Panel Data Model

We construct a panel data set such that each unit of observation is a member and each time period is one calendar month. The data set therefore contains monthly observations about each user, including the number of subscribers or followers that the user had in each month, his or her activities such as number of product reviews written in that month, the mean and variance of these ratings, and so on.⁸ We first test for serial correlations using the method suggested by Wooldridge (2002) and implemented in Stata package XTSMIAL (Drukker 2003). We find that for the number of ratings and number of nonrating articles, there is statistically significant first-order serial correlation ($p < 0.01$ and $p < 0.03$, respectively), and it remains even if we use a binary indicator for writing any reviews (versus none at all) or nonreview articles (both $p < 0.01$). We therefore turn to dynamic panel data models for these two outcome variables and use the Arellano-Bond estimator (Arellano and Bond 1991, Ghose 2009) to estimate the effect of incoming ties on volume of reviews and nonreview articles. This is estimated using the XTABOND procedure in Stata. To reduce skewness of data, we take natural logarithms of all count data before including them in the estimation.⁹ These include the number of reviews, number of nonreview articles, number of nonreciprocated incoming ties, number of nonreciprocated outgoing ties, and the number of reciprocated ties.¹⁰ As a robustness test, we also examined the

binary outcome variable for generating any product ratings or nonratings articles, respectively, in addition to the count measures. This specification is estimated using fixed-effect logit models. For the robustness test where we use the original scale for the number of ratings (not log-transformed), we estimate a negative binomial model to account for overdispersion, as well as zero-inflated Poisson models.

The other two outcome variables (the average and standard deviation of numeric ratings provided by the user) do not suffer from first-order serial correlation, as suggested by XTSMIAL ($p > 0.65$ and $p > 0.78$, respectively). However, there is a potential selection bias: There are no observations on the mean and standard deviations of ratings unless there is at least one rating. We therefore estimate a Heckman selection model with user fixed-effects, using the two-step procedure. Furthermore, it may not be meaningful to calculate the mean and standard deviation of ratings when there is only one rating in that month. As a result, we also examine the robustness of results using different minimum numbers of reviews to calculate mean and standard error.¹¹ More formally, the models that we estimate are as follows:

$$\begin{aligned} NbrReviews_{it} &= \alpha_0^1 + \alpha_1^1 NbrReviews_{i(t-1)} + \alpha_2^1 IN_{it} + \alpha_3^1 OUT_{it} \\ &\quad + \alpha_4^1 RCP_{it} + \alpha_5^1 IN_{it}^2 + \alpha_6^1 OUT_{it}^2 + \alpha_7^1 RCP_{it}^2 \\ &\quad + \alpha_8^1 Tenure_{it} + \nu_i^1 + \epsilon_{it}^1 \end{aligned} \quad (1)$$

$$\begin{aligned} NbrArticles_{it} &= \alpha_0^2 + \alpha_1^2 NbrArticles_{i(t-1)} + \alpha_2^2 IN_{it} + \alpha_3^2 OUT_{it} \\ &\quad + \alpha_4^2 RCP_{it} + \alpha_5^2 IN_{it}^2 + \alpha_6^2 OUT_{it}^2 + \alpha_7^2 RCP_{it}^2 \\ &\quad + \alpha_8^2 Tenure_{it} + \nu_i^2 + \epsilon_{it}^2 \end{aligned} \quad (2)$$

$$\begin{aligned} Valence_{it} | (NbrReviews_{it} \geq k_1) &= \alpha_0^3 + \alpha_1^3 IN_{it} + \alpha_2^3 OUT_{it} + \alpha_3^3 RCP_{it} + \alpha_4^3 IN_{it}^2 + \alpha_5^3 OUT_{it}^2 \\ &\quad + \alpha_6^3 RCP_{it}^2 + \alpha_7^3 Tenure_{it} + \nu_i^3 + \lambda_3(\cdot) + \epsilon_{it}^3 \end{aligned} \quad (3)$$

$$\begin{aligned} Variance_{it} | (NbrReviews_{it} \geq k_2) &= \alpha_0^4 + \alpha_1^4 IN_{it} + \alpha_2^4 OUT_{it} + \alpha_3^4 RCP_{it} + \alpha_4^4 IN_{it}^2 + \alpha_5^4 OUT_{it}^2 \\ &\quad + \alpha_6^4 RCP_{it}^2 + \alpha_7^4 Tenure_{it} + \nu_i^4 + \lambda_4(\cdot) + \epsilon_{it}^4 \end{aligned} \quad (4)$$

The left-hand side of the above models refers to the dependent variables that we described earlier, i.e., number of ratings and nonrating articles written in each month and the average and variance of the ratings provided. k_1 in Equation (3) and k_2 in Equation (4) are integers and refer to the threshold that we

⁸ For the main analysis and its results, we estimate the panel data model using data from January 2007 to July 2009. Our results are not sensitive to the choice of starting points (e.g., January 2005, January 2006, or July 2008).

⁹ We obtain consistent results when count variables such as in-degree and number of reviews are included in the estimation without the log transformations, as reported in Table 4. Inactive members (no contributions in terms of reviews and articles, and no new activities in the web-of-trust) are not included in the estimate because they do not provide meaningful variations in either the independent variable or dependent variable.

¹⁰ In the reported results we differentiate between pure in-degrees versus reciprocated ties to recognize that a small portion of the ties may be reciprocated. For instance, John may have 20 incoming ties, out of which five are from members that John also trusts. We hence separate the five reciprocated ties from the remaining "pure" in-degree of 15. Results are, however, not sensitive to this; we obtain the same results if we ignore the potential overlap or reciprocation of ties.

¹¹ For example, we tested defining "selection" as writing at least three ratings so that the mean and standard deviation of ratings will only be calculated when there are at least three ratings from that user in that month.

use to calculate mean or variance of the ratings generated, with $k_1 \geq 1$ and $k_2 \geq 2$. *IN*, *OUT*, and *RCP* refer to the number of nonreciprocated (pure) incoming ties, nonreciprocated (pure) outgoing ties, and reciprocated ties of the user as of the first day of that month. *TENURE* refers to the number of months that the user had been on the site. ν are individual fixed-effects, and ϵ are the error terms. $\lambda(\cdot)$ in Equations (3) and (4) refer to the inverse mills ratio from the first stage of the Heckman selection model. In one of our robustness tests, we include a quadratic term for the number of months on site to allow for nonlinear time effects, and results are highly consistent.

In the specifications above, we do not control for existing product reviews, yet the literature suggests that they are likely to affect the users' behaviors as well (Moe and Trusov 2011). Controlling for existing product reviews in our model however is challenging because of the imperfect matching between our level of analysis and the data on existing product ratings. Specifically, in our model, each observation corresponds to a user-month pair. In a given month, a user may review multiple products. Some of those products may have zero ratings, whereas others may have several. To control for existing product reviews, we aggregate the existing ratings of the products that the user reviewed in that month by creating the following variables: (1) the average of those products' number of reviews up to the month in consideration; (2) the average of those products' average ratings up to the month in consideration; and (3) the average of the standard deviation of those products' ratings, also up to the month in consideration. We then incrementally add these variables to the panel data model as additional controls. An important downside to this approach however, is that such an aggregation process is inherently imprecise. For that reason, we retain the models described previously (without these variables) as our main specification but report the results of these models when controlling for these three variables in an online companion to this paper (available at <http://dx.doi.org/10.1287/isre.2013.0512>). Those results are qualitatively consistent with what we report in §5.4.

5.3. Matching

To test the robustness of the results obtained from the panel data models, we further take an entirely orthogonal approach and consider matching instead (Heckman et al. 1998). Matching methods are increasingly popular among empirical researchers in IS, and for our study, the matching method helps overcome two limitations of the panel data method: (1) it may be arbitrary to use calendar months as period cutoffs, and (2) it does not consider the intensity in which a given number of incoming ties arrive. For instance,

receiving 10 incoming ties over two days should have a different influence on user behavior than receiving 10 incoming ties over 10 days, even though the number of additional incoming ties is the same. We refer to this as levels of *treatment strength*. Specifically, we define treatment as receiving X incoming ties over Y days. A larger X/Y ratio indicates higher strength of treatment, whereas a lower ratio indicates a lower strength of treatment. If A receives three incoming ties in one day, and B receives three incoming ties over three days, then the effect on behavior should be stronger for A than for B. By using different combinations of X and Y , we can better understand effects of various treatment strengths.¹²

Suppose $X = 3$ and $Y = 2$; we apply the matching method in the following manner. We are interested in how a review writer's behavior changes after he or she receives three incoming ties in two consecutive days. Since we have the time stamp of all incoming ties, we identify all cases in our data where a user receives three incoming ties in two consecutive days. At each occurrence, the other users who were not "treated" are potential matches. Among these potential matches, we use normalized Euclidean distance to identify one user¹³ who is most similar to the treated user on user-level characteristics, including pure in-degree, pure out-degree, reciprocated degree, existing number of reviews, and number of months on site. Once matched pairs are identified, we identify their activities in the 30 days¹⁴ before and after that and then compare them in a difference-in-differences manner to estimate the effect. Formally, using y to indicate the outcome variable, subscripts 0 (before) and 1 (after) to indicate the time, and subscripts t (treated) and m (match) to indicate the users, our estimate¹⁵ is equal to $E[(y_{t1} - y_{t0}) - (y_{m1} - y_{m0})]$. We test our hypotheses using several combinations of X and Y and examine the robustness of the findings across these combinations as well as against those from the dynamic panel data models mentioned above.

It should be noted that despite its popularity among empirical researchers, the matching method is

¹² One limitation of this approach is that it does not allow direct tests of second-order effects.

¹³ We use normalized Euclidean distance and one-on-one matching to ensure their computational efficiency and tractability. If the matched units' distance is more than three times the standard deviation of all distances among matched pairs, those matches are dropped. This threshold is consistent with Aral et al. (2009).

¹⁴ We use 30 days to retain consistency with the monthly aggregation in our panel data models, but the results are not sensitive when we change the window to 45 days or 60 days.

¹⁵ For valence of ratings, this estimate is conditional on both the treated and matched units having provided at least one both before and after the treatment; for variance, the minimum is two.

Table 3 Panel Data Models of Reviews Occurrence (Log Scale)

	Arellano-Bond estimate for number of ratings (log)	Arellano-Bond estimate for number of ratings (log), no outliers	Fixed-effect logit model, for 1 (provide ratings)
Pure in-degree (natural log)	0.642*** (0.025)	0.383*** (0.024)	0.532*** (0.046)
Pure out-degree (natural log)	0.215*** (0.027)	0.207*** (0.026)	0.016 (0.050)
Reciprocated ties (natural log)	0.775*** (0.031)	0.329*** (0.030)	0.438*** (0.053)
Pure in-degree ²	−0.374*** (0.010)	−0.236*** (0.010)	−0.028*** (0.009)
Pure out-degree ²	0.046*** (0.009)	−0.003 (0.009)	−0.044*** (0.012)
Reciprocated ties ²	−0.257*** (0.009)	−0.119*** (0.009)	−0.015 (0.011)
Log number of ratings ($t - 1$)	0.244*** (0.003)	0.195*** (0.003)	(N/A) (N/A)
Number of months on site (natural log)	−0.587*** (0.009)	−0.532*** (0.008)	−1.460*** (0.026)
Intercept	2.701*** (0.035)	2.327*** (0.036)	1.361*** (0.082)
<i>N</i>	138,476	135,153	161,182

Notes. This table reports results on the volume of ratings using panel data. The first column reports the results from the Arellano-Bond linear dynamic panel data model, where the dependent variable is the log number of ratings provided by a user in a month, the main independent variable is the number of incoming ties (pure in-degree), and independent variables are log-scaled to reduce skewness. Results are consistent when raw metrics are used (see Table 4). The second column results are derived from the same model except that outliers and influential observations are removed. The third column reports the results from a fixed-effect logit model, where the binary dependent variable is whether the user provided any ratings in that month. All three results suggest that a higher number of incoming ties is associated with a higher probability of providing ratings and also a higher number of ratings. Results remain consistent when we do not distinguish between pure incoming ties and reciprocated ties. Robust standard errors are reported in parentheses under coefficients.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

not without its own limitations. An important weakness is that matches are constructed only according to observable variables, and hidden bias needs to be addressed. In our context, though, this concern may be alleviated. First, since the treatment that we are interested in is incoming ties from peer users of the site who are typically *silent strangers* to those being followed, the person being followed is unlikely to possess unobserved information about the follower that we cannot observe as researchers.¹⁶ Second, *in equilibrium*, network position of a user on the WOT—as indicated by the in-degree and out-degree metrics used in the matching process—should already capture most, if not all, unobservable characteristics of that user. As a basic rationale in sociological and economic studies of networks, a person's characteristics and actions should determine his or her position in a social network. In other words, unobservable factors that may simultaneously affect a user's popularity and his or her behaviors are likely subsumed in their network metrics. Conditional on these observable network metrics, therefore, the treatment (receiv-

ing new incoming ties) can be considered exogenous. Nonetheless, readers should still use caution when interpreting the results or generalizing them into other contexts.

5.4. Discussion of Results

Our results from the dynamic panel data as well as the matching method are highly consistent. Tables 3 through 8 report the results of various panel data specifications for the four outcomes of interest. It should be noted that different columns may contain estimates from different empirical models and dependent variables.¹⁷ Table 9 provides the results from the matching method.

Volume of Ratings and Nonrating Reviews.

Results from the panel data models lend support to H1: the coefficient on the number of nonreciprocated incoming ties (pure in-degree) is positive and statistically significant, whereas the coefficient on the quadratic term is negative and statistically significant. These results are consistent across multiple specifications for the log number of ratings. It is also

¹⁶ For dyads of nonstrangers on WOT, they are likely to be connected by reciprocated ties rather than pure incoming ties. The reciprocated ties are controlled for in all our models.

¹⁷ To conserve space, we do not report results from the first stage (i.e., "selection") of the Heckman models because they are highly consistent with those in the final column of Table 3.

Table 4 Panel Data Models of Reviews Occurrence (Original Scale)

	Arellano-Bond estimate for number of ratings	Fixed-effect logit model for 1 (provide ratings)
Pure in-degree	6.795e-03*** (3.101e-04)	2.285e-02*** (1.414e-03)
Pure out-degree	0.002*** (0.001)	0.012*** (0.002)
Reciprocated ties	0.009*** (0.001)	0.008*** (0.002)
Pure in-degree ²	-5.885e-06*** (3.721e-07)	-2.024e-05*** (1.763e-06)
Pure out-degree ²	-3.932e-06*** (1.246e-06)	-1.265e-05*** (3.717e-06)
Reciprocated ties ²	-1.663e-05*** (1.282e-06)	-2.528e-05*** (4.174e-06)
Number of months on site	-0.027*** (0.001)	-0.104*** (0.002)
Intercept	-1.147*** (0.032)	1.660*** (0.110)
<i>N</i>	138,476	161,182

Notes. This table reports robustness-test results on the occurrence or incidence of ratings using a panel data setup with variables in original scale (not log-transformed). The first column reports the results from the Arellano-Bond linear dynamic panel data model, where the dependent variable is the number of ratings provided by a user in a month and the main independent variable is the number of incoming ties (pure in-degree). The second column reports the results from a fixed-effect logit model, where the binary dependent variable is whether the user provided any ratings in that month. Results confirm that a higher number of incoming ties (pure in-degree) is associated with a higher probability of providing ratings and also a higher number of ratings. These results remain consistent when we do not distinguish between pure incoming ties and reciprocated ties. Robust standard errors are reported in parentheses under coefficients.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table 5 Heckman Model Results for the Valence (Mean) of Ratings Provided

	"Selection" defined as providing one rating or more	"Selection" defined as providing two ratings or more	"Selection" defined as providing three ratings or more
Pure in-degree	-0.079*** (0.026)	-0.112*** (0.024)	-0.112*** (0.024)
Pure out-degree	0.024 (0.023)	0.032 (0.023)	0.032 (0.023)
Reciprocated ties	0.042* (0.022)	-0.030 (0.024)	-0.030 (0.024)
Pure in-degree ²	0.006* (0.003)	0.009*** (0.003)	0.009*** (0.003)
Pure out-degree ²	0.007 (0.005)	0.008 (0.005)	0.008 (0.005)
Reciprocated ties ²	-0.017*** (0.005)	-0.002 (0.004)	-0.002 (0.004)
Time on site	0.062** (0.031)	0.053* (0.029)	0.053* (0.029)
Intercept	4.348*** (0.096)	4.238*** (0.086)	4.247*** (0.090)
<i>N</i>	56,127	49,632	47,413

Notes. This table reports the results of Heckman models (second stage) for the valence (mean) of ratings provided by each user in each month, since valence is only defined when the user has written something in each month. Independent variables are log-transformed. Three columns use different thresholds to calculate the valence of ratings: The first is the simple average; the second only calculates if there are two or more ratings provided in that month; and the third only calculates if there are three or more ratings provided in that month. The selection stage results of the Heckman model are not reported for brevity and also because they are consistent with the final column in Table 3. More incoming ties are associated with a lower average of ratings, regardless of the threshold for valence calculation. Robust standard errors are reported in parentheses under coefficients.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table 6 Heckman Model Results for the Standard Deviation of Ratings Provided

	Selection defined as providing two or more ratings	Selection defined as providing three or more ratings
Pure in-degree	0.238** (0.113)	0.209* (0.110)
Pure out-degree	−0.007 (0.024)	−0.039* (0.022)
Reciprocated ties	0.164** (0.069)	0.162** (0.077)
Pure in-degree ²	−0.018* (0.010)	−0.017* (0.010)
Pure out-degree ²	−0.019* (0.010)	−0.010 (0.009)
Reciprocated ties ²	0.016 (0.010)	0.007 (0.008)
Time on site	−0.447** (0.201)	−0.355* (0.194)
Intercept	−0.463 (0.644)	−0.137 (0.633)
<i>N</i>	49,632	47,413

Notes. This table reports the results of Heckman models for the standard deviation of ratings provided by each user in each month since standard deviation is only defined when the user has written something in each month. Independent variables are log-transformed. The two columns use different thresholds to calculate the standard deviation of ratings: the first one calculates standard deviation only if there are two or more ratings provided in that month; the second only calculates if there are three or more ratings provided in that month. The selection stage results of the Heckman model are not reported for brevity and also because they are consistent with the final column in Table 3. More incoming ties are associated with higher standard deviations of ratings, regardless of the threshold for calculation (though sample sizes for the outcome stage estimate will be smaller). Robust standard errors are reported in parentheses under coefficients.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

qualitatively consistent from the count data model of the number of ratings (original scale) as well as a binary outcome model for whether any reviews were written. We also note that these results remain consistent even if we do not differentiate between reciprocated versus nonreciprocated ties between the writer and the reader. Hence, more incoming ties result in more contribution from the user; however, the marginal effect is decreasing. Results from matching provide more straightforward interpretations of the effect's magnitude. For instance, receiving three incoming ties in three consecutive days can, on average, motivate the user to provide 6.6 more new product ratings in the subsequent 30 days and write one more nonrating article. The magnitude of this effect varies by the intensity of the treatment: if the number of incoming ties is spread out over six days instead of three (a weaker treatment), the increase is about four. In unreported tests, we also find that the total number of *words* written by these members increases as well. These results show that there is indeed a robust

encouragement effect when the writer becomes more popular, although the effect is stronger for those who have a smaller audience than those with a larger audience. In other words, to encourage the sharing of consumption experiences, and if the emphasis is on the number of reviews or nonreview articles, website administrators may be better off showcasing the reviews written by "up-and-coming" contributors, so they have a higher chance of receiving new incoming ties, rather than established "celebrities." Given that writing product reviews is largely a public good (Chen et al. 2010), these results are particularly important for social media sites.

Valence of Ratings. Our results from the Heckman models provide support for H2: the coefficient on the number of incoming ties is negative and statistically significant, whereas the coefficient on the quadratic term is positive and statistically significant. We obtained consistent results when we increase the threshold for the calculation of valence, i.e., if we do not consider the average meaningful when the user has written at least two or three ratings that month. These results suggest that, all else equal, as more peers trust the user, he or she is indeed likely to write more negatively. The marginal effect is decreasing, though this effect is weaker. Easier interpretations come from matching: Receiving three incoming ties in two days, on average, is followed by the user providing ratings about −0.03 stars lower.

At first look, a decrease of 0.03 stars seems rather miniscule. However we note that this is out of a small range of five stars. To better understand this effect size, we can think of this number in several different ways. First, at the rate of three new ties every two days, if this "trend" continues up to 60 days (multiplied by 30), the cumulative effect can be one fewer star out of five (by design, ratings on epinions.com are between one and five stars). Second, epinions.com only allows discrete ratings (one star, two stars, etc.), so even a small decrease in the users' "latent" rating could potentially translate into a one-star difference in the actual rating that he or she provides. This could be especially detrimental if the writer is the first person to review a product. Third, this effect is the result of only three incoming ties; for users who have a large following (e.g., dozens or hundreds of followers), it can potentially translate into a significant difference. Given the potential number of new members who join the site over time—who will initiate those incoming ties after they join—cumulatively, this effect can be quite significant. And finally, this effect may be further magnified because of the increasing interactions among expressed opinions (Moe and Trusov 2011). Therefore, although effect size may be ultimately a

Table 7 Panel Data Models for Incidence of Nonreview Articles (Log Scale)

	Arellano-Bond estimate for number of nonrating articles (log)	Arellano-Bond estimate for number of nonrating articles (log, no outliers)	Fixed-effect logit model for 1 (write any nonrating articles)
Pure in-degree	0.104*** (0.006)	0.019*** (0.005)	0.859*** (0.168)
Pure out-degree	0.000 (0.007)	−0.013** (0.006)	0.744*** (0.191)
Reciprocated ties	0.111*** (0.008)	0.036*** (0.007)	2.574*** (0.227)
Pure in-degree ²	−0.077*** (0.003)	−0.016*** (0.002)	−0.083*** (0.026)
Pure out-degree ²	0.012*** (0.002)	0.014*** (0.002)	−0.160*** (0.037)
Reciprocated ties ²	−0.050*** (0.002)	−0.017*** (0.002)	−0.254*** (0.037)
Number of nonrating articles, lagged	0.252*** (0.003)	0.125*** (0.003)	(N/A) (N/A)
Time on site	−0.037*** (0.002)	−0.019*** (0.002)	−1.979*** (0.096)
Intercept	0.346*** (0.010)	0.104*** (0.009)	−5.612*** (0.251)
<i>N</i>	138,476	135,153	161,182

Notes. This table reports results on the number of nonrating articles using a panel data with independent variables log-transformed (results are consistent when using raw scale; not reported for brevity). The first column reports the results from the Arellano-Bond linear dynamic panel-data model, where the dependent variable is the log number of nonrating articles written by a user in a month. The second column results are from the same model except that outliers and influential observations are removed. The third column reports the results from a fixed-effect logit model, where the binary dependent variable is whether the user provided any ratings in that month. All three results suggest that a higher number of incoming ties is associated with a higher probability of providing nonrating articles and also a higher number of such articles. Results are consistent when we do not distinguish between pure incoming ties and reciprocated ties. Robust standard errors are reported in parentheses under coefficients.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table 8 Results for Text Features

	Emotional words	Readability: Gunning-Fog Index	Readability: lexical density
Pure in-degree (natural log)	−0.297*** (0.041)	0.853*** (0.330)	−1.076*** (0.084)
Pure out-degree (natural log)	0.100** (0.040)	−0.158 (0.321)	0.213*** (0.081)
Reciprocated degree (natural log)	0.281*** (0.045)	−2.036*** (0.358)	0.612*** (0.091)
Pure in-degree ²	−0.033*** (0.007)	0.185*** (0.054)	−0.123*** (0.014)
Pure out-degree ²	−0.037*** (0.010)	0.074 (0.076)	−0.121*** (0.019)
Reciprocated degree ²	−0.024*** (0.009)	0.204*** (0.069)	0.001 (0.018)
Number of months on site (natural log)	−0.424*** (0.032)	−1.018*** (0.258)	−0.956*** (0.066)
Intercept	2.873*** (0.205)	27.150*** (1.636)	3.652*** (0.419)
<i>N</i>	56,127	56,127	56,127

Notes. This table reports the results of Heckman models (second stage) for text features of reviews (aggregated) provided by each user in each month. Independent variables are log-transformed. Three columns report results on different outcome variables: The first is the portion of emotional words; the second, readability of the texts measured by the Gunning-Fog Index; and the third, readability measured by lexical density. The selection stage results of the Heckman model are not reported for brevity and also because they are consistent with the final column in Table 3. More incoming ties are associated with fewer emotional words in the texts and also increased readability of the texts. See paper for detailed discussions on the practical significance of these results. Robust standard errors are reported in parentheses under coefficients.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table 9 Robustness: Effect of Receiving X Ties Over Y Days on Ratings and Nonrating Articles

X	Y	$\Delta\Delta$ (reviews written)	$\Delta\Delta$ (mean of ratings)	$\Delta\Delta$ (standard deviation of ratings)	$\Delta\Delta$ (nonreview articles written)
3	2	7.0**	−0.03***	0.17**	1.02**
3	3	6.6***	−0.03***	0.16***	0.8***
3	4	3.99***	−0.02***	0.12***	1.47**
3	5	4.12***	−0.01**	0.09***	1.21***
3	6	3.9***	−0.01**	0.08***	1.1***
4	3	7.66***	−0.06**	0.17***	1.23**
4	4	7.28***	−0.04***	0.18***	1.06***
4	5	6.76***	−0.01***	0.17***	0.89***
4	6	6.44***	−0.01**	0.15***	0.74***
5	4	8.27**	−0.09***	0.18**	1.3**
5	5	7.9***	−0.05***	0.2***	1.24***
5	6	7.38***	−0.04***	0.19***	1.14***

Notes. This table reports the effect of receiving X incoming ties in Y consecutive days and measuring the effect over a 30 day period, using the matching method discussed in §5.3 of the paper. Results from various combinations of X and Y are reported and are qualitatively consistent with those from the panel data estimates.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table 10 Effect of Receiving Three Incoming Ties Over Two Days on Linguistic Features of Reviews

Linguistic feature	Average treatment effect (ATE)	Standard error	t -value	Lower bound of 95% confidence interval of treatment effect	Upper bound of 95% confidence interval of treatment effect
Emotional words	−1.092	0.030	−36.637	−1.151	−1.034
Positive emotion words	−0.763	0.021	−36.382	−0.804	−0.722
Negative emotion words	−0.321	0.011	−30.442	−0.341	−0.300
Readability: Gunning-Fog Index	0.192	0.054	3.517	0.085	0.298
Readability: lexical density	−7.877	0.256	−30.813	−8.379	−7.376

Notes. This table reports the effect of receiving three incoming ties in two consecutive days on the linguistic features of product reviews that users generate. The methods are described in §5.3 of the paper. Emotional words (positive and negative) and lexical density are measured as percentages, whereas the Gunning-Fog Index is interpreted as the number of years of education required for understanding the text.

matter of judgment, the aggregate effect from multiple opinion writers and their cumulative effect over time could be quite substantial.

We further conduct an additional test using an alternative measure of valence. Specifically, instead of the mean, we look at the number of m -star ratings that a user writes in a given month, where $m = 1, 2, 3, 4$, or 5 . This provides us a more concrete measure of the relative frequency that each rating is provided by the user. Our results show that although all five types of ratings increase in their frequency (consistent with H1), the incidence of lower ratings occurs more frequently than higher ones, consistent with the finding on valence.

Variance of Ratings. Our results show that the coefficient on the number of incoming ties (pure in-degree) is positive and statistically significant, whereas the coefficient on its quadratic term is negative and statistically significant. These results are also not sensitive to the threshold for calculating standard deviations. From the matching results, we see that for the range of X - Y combinations that we examine, the increase in the standard deviation of ratings range between 0.08 and 0.2. Since the average user has a standard deviation of 0.46 in his or her ratings, this

effect is not trivial. In other words, users who are more trusted by others are more likely to express a wider range of opinions. Prior studies have shown that product reviews tend to be bimodal, i.e., either extremely positive or extremely negative (Hu et al. 2006). Our results indicate that user interactions, such as the opportunity to subscribe to each other, may further encourage divergent views on products. On the other hand, as we discussed in the hypothesis section, because of the range of star ratings, such increases in variations is unlikely to be monotonic, and marginal effects will be decreasing, as suggested by the coefficient of the quadratic terms.

Text Feature of Reviews. Readers of product reviews don't just count the number of stars; they read them. Recent papers have shown the informational value of such textual features of product reviews (Butler and Wang 2012, Pavlou and Dimoka 2006). Hence, our understanding of how subscribers affect writer behavior will not be complete without looking at the linguistic characteristics of their writing, or the way they write. Using both the regression model and matching methods, we next study the effect of an increasing audience on writing features, especially the readability and emotional level of the texts (H4).

To begin with, reviews should be *readable* and easy to understand to sway consumer decisions. To examine the readability of reviews, we use two popular readability metrics: the lexical density metric (Keegan and Kabanoff 2008, Read 2000) and the Gunning-Fog Index (Gunning 1969, Kasper and Morris 1988, Sawyer et al. 2008, Teichroew et al. 1967). Their formulae are as follows, respectively:

$$\text{Lexical Density}^{18} = \left(\frac{\text{Number of Unique Words}}{\text{Number of Words}} \right) \times 100 \quad (5)$$

$$\text{Gunning-Fog Index}^{19} = 0.4 \times \left(\frac{\text{Number of Words}}{\text{Number of Sentences}} + \frac{\text{Number of hard words}}{\text{Number of Words}} \times 100 \right) \quad (6)$$

Here, "hard words" are defined as words with three syllables or more.²⁰ Lexical density (LD) measures the degree of information contained in texts. Higher density suggests that a text contains more information and is more difficult to read. The Gunning-Fog Index, on the other hand, measures the estimated number of years of education that a reader must have to understand the text. Texts with a higher Gunning-Fog Index are more difficult to understand.

We further examine the emotional content of reviews. Text sentiments are not only important complements to the valence of numeric ratings, they are also more granular and will also reflect the objectivity of reviews. All else equal, a review that uses more emotional words (either positive or negative) is less likely to be objective.²¹ To quantify the extent of emotional words usage in the texts (the opposite of objectivity), we use the Linguistic Inquiry and Word Count package (LIWC; Pennebaker et al. 2006), which has been extensively used in published studies in management and other fields (Bednar 2012, Berger and

Milkman 2012, Brett et al. 2007). In addition to a total count of emotional words in a text, LIWC further measures the extent of positive emotions as well as negative emotions.

We calculated and used these metrics as dependent variables in the panel data model and the matching process by aggregating all reviews a user writes in a given calendar month (using the panel data method described in §5.2) or in the 30 days after receiving "treatment" (using the matching method described in §5.3). Some interesting findings emerge from the results, as reported in Tables 8 through 10. As users are trusted by more of their peers, they are less likely to use emotional words (1.8% of all words for treated units versus 2.9% of all words for untreated matches; the difference of 1.092% is reported in the first data column of Table 9), regardless of whether the emotion is positive (1.23% versus 1.99%) or negative (0.51% versus 0.83%). These suggest that users who are trusted by more peers tend to become more objective in their reviews; they increasingly sound like an authority in their writings, confirming our hypothesis.

By contrast, there seems to be only modest effect on the reviews' readability metrics. As a user gathers a larger following, the lexical density of his reviews decreases (32% versus 40%), whereas the Gunning-Fog Index increases (9.5 versus 9.3). However, the difference in Gunning-Fog Index only amounts to fewer than 3 months of education (12 months per year $\times (9.5 - 9.3)$), and lexical density scores below 40% are all considered low (see footnote 18). Therefore, regardless of the sign of the coefficients, these results suggest that readability of the reviews appears to be quite stable even if the writer gathers a large online audience. It is the words being used (emotional or nonemotional) that show a more interesting pattern. Nonetheless, linguistic analysis is still an active field of research, and we may derive new insights as new tools and methods emerge.

Control Variables. Before we conclude this section, we discuss some findings on the control variables. A user's number of outgoing ties (pure out-degrees), for instance, is fully controlled by the user. It may be because of this endogeneity that results on this variable are quite sensitive to model specification changes. By contrast, the number of reciprocated ties is also positively associated with the number of ratings and nonrating articles, suggesting that reciprocated ties—which may indicate deeper connections, acquaintances, or friends—may also motivate users to contribute. However, mutual connections may suggest that there is information about the users themselves that is not observable to us as researchers. Therefore, effects of the reciprocated ties should only be interpreted as correlational. In addition, holding

¹⁸ Texts with an LD of 60%–70% are considered dense, and those between 40% and 50% are considered not dense (source: <http://www.usingenglish.com/glossary/lexical-density-test.html>, accessed July 8th, 2013).

¹⁹ Source: <http://www.usingenglish.com/glossary/lexical-density-test.html>, accessed July 8th, 2013.

²⁰ While we coded these measures in our program, the results are highly consistent with those produced by several websites such as <http://www.usingenglish.com/resources/text-statistics.php> (accessed July 8th, 2013).

²¹ Objectivity of text is an active area of research in machine learning and is beyond the scope of our paper. It is especially difficult when multiple domains are involved (e.g., Lu et al. 2013 focus on movie ratings). We therefore indirectly assess the objectivity of texts by studying the portion of emotional words identified via LIWC.

the incoming and outgoing ties constant, the longer the user is on the site (time since registration), the fewer ratings and nonrating articles that they contribute. This is also reasonable: For two users who have similar local network structures in WOT (in-degrees and out-degrees), the person who had been on the site longer is the one who was only able to gather the same number of followers over a longer period of time—the "treatment intensity" is obviously lower. Interestingly, the effect of time on the valence of reviews is negative. It appears that writers who are on the site longer but do not observe their readership growing are likely to be increasingly critical. It may be due to their frustration, or it may be an effort to increase their audience by acting negatively (Amabile 1983); differentiating these explanations can be an interesting area for future research.

6. Implications and Future Research

A direct implication of our findings is that online user interactions, especially in the form of subscriptions, that have become ubiquitous in social media do affect user behavior. For websites that are trying to increase the sheer volume of activity on their sites (i.e., website traffic), our results confirm that online interactions indeed encourage users to write more reviews and nonreview articles. The decreasing marginal effect (H1) suggests that user-generated content (UGC) websites' promotional efforts may achieve better results if spent on featuring up-and-coming content creators rather than celebrities so that they are more visible and more likely to receive incoming ties from other website users. In light of our results on the linguistic features of reviews, such efforts can induce more objective reviews as well, thereby providing more value to website users in general. Meanwhile, these incoming ties also affect the valence and variance of reviews being generated. Marketers and website owners should carefully weigh these trade-offs when they manage product review platforms. For design scientists and website administrators, our results also suggest that such interactions should be taken into account when aggregating the opinion of the crowd. Unlike well-administered survey studies that ensure each participant answers independently, online product reviews reflect the interaction among expressed opinions (Moe and Trusov 2011) as well as interactions among users. To our knowledge, our paper is the first to document the latter effect.

Our study contributes to several streams of literature, especially the ones that we draw on for our hypotheses. It extends prior research on peer effect and social influence and identifies a new mechanism through which "silent" peers, who often constitute the majority of website users, may influence the

behavior of others by simply providing attentive ears. It explores new dynamics in the opinion leadership literature and provides some preliminary evidence that the presence and intensity of opinion *followers* may also influence the behaviors of opinion leaders. It also supplements the large literature on online communities (Ren et al. 2012), especially how user interactions affect participation, and provides more granularities regarding how users react to their interactions with peers of an online community.

Our paper further contributes to the ever-growing literature on online word-of-mouth (Dellarocas 2003, Forman et al. 2008) in at least three aspects. First, we provide new evidence on how the *generation* of online product reviews is related to the presence of peers, even if those peers are silent and do not express their opinions themselves. Second, ours is one of the first studies to look at online user reviews from the perspective of *writers*, whereas prior research has largely examined reviews from the product's perspective (Liu 2006) or how ratings may influence each other (Moe and Trusov 2011). Third, in addition to the well-established metrics of volume and valence, we build upon and extend studies that examine the textual content and linguistic features of user reviews (Lu et al. 2013, Pavlou and Dimoka 2006), measuring both the emotional content and readability of reviews.

Our results also address the broader "wisdom of crowd" (Ransbotham et al. 2012) phenomenon. User reviews are often cited as an example of this phenomenon, in the sense that even if one consumer's opinion is idiosyncratic, aggregating opinions from many consumers can help measure the product's objective quality. Interestingly, lab studies (Lorenz et al. 2011) show that social interactions and social influence can undermine this "wisdom of crowd" effect. Our study offers complementary field evidence that users' product review behaviors change when there is an increasingly larger followership, even if the followers are silent—which is typically not possible in labs.

The current study can be extended in several directions. For instance in our data set, we do not observe the actual adoption or purchase behavior of the review writers. It is possible that product reviewers' choice of products may also change over time as they become increasingly trusted by peers. Such information will help explain one of the patterns that we observed in the data: as they garner a larger readership, review writers tend to become increasingly negative. If we have information about their purchase behavior over time, we will be able to answer what was causing that change. There are at least three possibilities. First, they may change products that they buy, especially in terms of the products' life cycle; the users may become "innovators" in the taxonomy

of Rogers' (1995) product diffusion process. Second, they may continue to buy the same products but just become more stringent and more difficult to please. And third, they may consume the same products and retain their level of stringency but may modify the subset of opinions that they choose to reveal online. This has important implications for website designers and also marketing practitioners but is well beyond the scope of our current data set. Controlled lab experiments should be better able to tease out these effects.

It will also be interesting and worthwhile to extend our framework to other contexts of UGC where users can subscribe to or follow each other and study how such interactions affect user behavior. To our knowledge, our study is one of the first to study the consequences of becoming "popular" in the generation of online product reviews, but many other UGC sites have similar subscription features that allow similar user interactions. Examples include Twitter (follow), Facebook (like), and social investment platforms such as Covestor.com. Although behavioral outcomes will differ from context to context, it is quite plausible that a larger audience will change the behavior of those being followed, such as their risk preference for the experts' stock choice on social investment platforms. These are fertile grounds for future research.

Supplemental Material

Supplemental material to this paper is available at <http://dx.doi.org/10.1287/isre.2013.0512>.

Acknowledgments

The authors thank participants at the 2012 INFORMS Annual Meeting and research workshops at Carnegie Mellon University, Temple University, Texas Tech University, University of Auckland, University of Arizona, University of Maryland, and University of Minnesota for their comments and suggestions. They are especially grateful for the constructive comments and insights from Ram Gopal (senior editor), Ming Fan (associate editor), and two anonymous reviewers. All errors remain their own.

References

- Adair JG (1984) The Hawthorne effect: A reconsideration of the methodological artifact. *J. Appl. Psych.* 69(2):334–335.
- Aggarwal R, Gopal R, Gupta A, Singh H (2012) Putting money where the mouths are: The relation between venture financing and electronic word-of-mouth. *Inform. Systems Res.* 23(3):976–992.
- Amabile TM (1983) Brilliant but cruel: Perceptions of negative evaluators. *J. Experiment. Soc. Psych.* 19(2):146–156.
- Aral S, Muchnik L, Sundararajan A (2009) Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proc. National Acad. Sci.* 106(51):21544–21549.
- Arellano M, Bond S (1991) Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Rev. Econom. Stud.* 58(2):277–297.
- Bateman PJ, Gray PH, Butler BS (2011) The impact of community commitment on participation in online communities. *Inform. Systems Res.* 22(4):841–854.
- Bednar MK (2012) Watchdog or lapdog? A behavioral view of the media as a corporate governance mechanism. *Acad. Management J.* 55(1):131–150.
- Berger J, Milkman KL (2012) What makes online content viral? *J. Marketing Res.* 49(2):192–205.
- Berger J, Sorensen AT, Rasmussen SJ (2010) Positive effects of negative publicity: When negative reviews increase sales. *Marketing Sci.* 29(5):815–827.
- Biddle BJ (1986) Recent developments in role theory. *Annual Rev. Sociology* 12:67–92.
- Bolton GE, Katok E, Ockenfels A (2004) How effective are electronic reputation mechanisms? An experimental investigation. *Management Sci.* 50(11):1587–1602.
- Brett JM, Olekalns M, Friedman R, Goates N, Anderson C, Lisco CC (2007) Sticks and stones: Language, face, and online dispute resolution. *Acad. Management J.* 50(1):85–99.
- Butler BS, Wang X (2012) The cross-purposes of cross-posting: Boundary reshaping behavior in online discussion communities. *Inform. Systems Res.* 23(3):993–1010.
- Cheema A, Kaikati AM (2010) The effect of need for uniqueness on word of mouth. *J. Marketing Res.* 47(3):553–563.
- Chen Y, Harper FM, Konstan J, Li X (2010) Social comparisons and contributions to online communities: A field experiment on Movielens. *Amer. Econom. Rev.* 100(4):1358–1398.
- Chintagunta PK, Gopinath S, Venkataraman S (2010) The effects of online user reviews on movie box office performance: Accounting for sequential rollout and aggregation across local markets. *Marketing Sci.* 29(5):944–957.
- Dellarocas C (2003) The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management Sci.* 49(10):1407–1424.
- Dellarocas C, Wood CA (2008) The sound of silence in online feedback: Estimating trading risks in the presence of reporting bias. *Management Sci.* 54(3):460–476.
- Dellarocas C, Gao GG, Narayan R (2010) Are consumers more likely to contribute online reviews for hit or niche products? *J. Management Inform. Systems* 27(2):127–157.
- Dellarocas C, Zhang X, Awad NF (2007) Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *J. Interactive Marketing* 21(4):23–45.
- Drukker DM (2003) Testing for serial correlation in linear panel-data models. *Stata J.* 3(2):168–177.
- Duan W, Gu B, Whinston AB (2008) Do online reviews matter? An empirical investigation of panel data. *Decision Support Systems* 45(4):1007–1016.
- Faraj S, Johnson SL (2011) Network exchange patterns in online communities. *Organ. Sci.* 22(6):1464–1480.
- Forman C, Ghose A, Wiesenfeld B (2008) Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Inform. Systems Res.* 19(3):291–313.
- Ghose A (2009) Internet exchanges for used goods: An empirical analysis of trade patterns and adverse selection 1. *MIS Quart.* 33(2):263–292.
- Gu B, Konana P, Rajagopalan B, Chen H-WM (2007) Competition among virtual communities and user valuation: The case of investing-related communities. *Inform. Systems Res.* 18(1):68–85.
- Gunning R (1969) The Fog index after twenty years. *J. Bus. Comm.* 6(2):3–13.
- Heckman JJ, Ichimura H, Todd P (1998) Matching as an econometric evaluation estimator. *Rev. Econom. Stud.* 65(2):261–294.

- Hu N, Pavlou PA, Zhang J (2006) Can online reviews reveal a product's true quality? Empirical findings and analytical modeling of online word-of-mouth communication. *Proc. 7th ACM Conf. Electronic Commerce* (ACM, New York), 324–330.
- Ibarra H, Andrews SB (1993) Power, social influence, and sense making: Effects of network centrality and proximity on employee perceptions. *Admin. Sci. Quart.* 38(2):277–303.
- Iyengar R, Van Den Bulte C, Valente TW (2011) Opinion leadership and social contagion in new product diffusion. *Marketing Sci.* 30(2):195–212.
- Kasper GM, Morris AH (1988) The effect of presentation media on recipient performance in text-based information systems. *J. Management Inform. Systems* 4(4):25–43.
- Keegan J, Kabanoff B (2008) Indirect industry- and subindustry-level managerial discretion measurement. *Organ. Res. Methods* 11(4):682–694.
- Li X, Hitt L (2008) Self selection and information role of online product reviews. *Inform. Systems Res.* 19(4):456–474.
- Liu Y (2006) Word of mouth for movies: Its dynamics and impact on box office revenue. *J. Marketing* 70(3):74–89.
- Lorenz J, Rauhut H, Schweitzer F, Helbing D (2011) How social influence can undermine the wisdom of the crowd effect. *Proc. National Acad. Sci.* 108(22):9020–9025.
- Lu Y, Jerath K, Singh PV (2013) The emergence of opinion leaders in a networked online community: A dyadic model with time dynamics and a heuristic for fast estimation. *Management Sci.* 59(8):1783–1799.
- Ma M, Agarwal R (2007) Through a glass darkly: Information technology design, identity verification, and knowledge contribution in online communities. *Inform. Systems Res.* 18(1):42–67.
- Marsden PV, Friedkin NE (1993) Network studies of social influence. *Sociol. Methods Res.* 22(1):127–151.
- Moe WW, Trusov M (2011) The value of social dynamics in online product ratings forums. *J. Marketing Res.* 48(3):444–456.
- Pavlou P, Dimoka A (2006) The nature and role of feedback text comments in online marketplaces: Implications for trust building, price premiums, and seller differentiation. *Inform. Systems Res.* 17(4):392–414.
- Pennebaker J, Francis M, Booth R (2001) *Linguistic Inquiry and Word Count: LIWC 2001* (Lawrence Erlbaum Associates, Mahwah, NJ), 71.
- Ransbotham S, Kane GC (2011) Membership turnover and collaboration success in online communities: Explaining rises and falls from grace in Wikipedia. *MIS Quart.* 35(3):613–627.
- Ransbotham S, Kane GC, Lurie NH (2012) Network characteristics and the value of collaborative user-generated content. *Marketing Sci.* 31(3):387–405.
- Read J (2000) *Assessing Vocabulary* (Cambridge University Press, Cambridge, UK).
- Ren Y, Harper FM, Drenner S, Terveen L, Kiesler S, Riedl J, Kraut RE (2012) Building member attachment in online communities: Applying theories of group identity and interpersonal bonds. *MIS Quart.* 36(3):841–864.
- Rice SC (2012) Reputation and uncertainty in online markets: An experimental study. *Inform. Systems Res.* 23(2):436–452.
- Rogers EM (1995) *Diffusion of Innovations* (Free Press, New York).
- Sawyer AG, Laran J, Xu J (2008) The readability of marketing journals: Are award-winning articles better written? *J. Marketing* 72(1):108–117.
- Shah PP (2000) Network destruction: The structural implications of downsizing. *Acad. Management J.* 43(1):101–112.
- Singh PV, Tan Y, Mookerjee V (2011) Network effects: The influence of structural capital on open source project success. *MIS Quart.* 35(4):813–829.
- Sohl JE (1999) The early-stage equity market in the USA. *Venture Capital: Internat. J. Entrepreneurial Finance* 1(2):101–120.
- Sorenson O, Stuart TE (2005) The evolution of venture capital investment networks. Working paper, Yale University, New Haven, CT.
- Sun M (2012) How does variance of product ratings matter? *Management Sci.* 58(4):696–707.
- Susarla A, Oh J-H, Tan Y (2012) Social networks and the diffusion of user-generated content: Evidence from YouTube. *Inform. Systems Res.* 23(1):23–41.
- Teichroew D, Joslin JC, Trieb SE, Maishshall WS, Kriebel CA, Randall DL (1967) Free for all. *Management Sci.* 13(6):378–382.
- Trusov M, Bodapati AV, Bucklin RE (2010) Determining influential users in Internet social networks. *J. Marketing Res.* 47(4):643–658.
- Wang C, Zhang M, Hann I-H (2010) Social influence in online product ratings. *Workshop on Information Systems and Economics (WISE), December 2010, St. Louis*.
- Wasko MM, Faraj S (2005) Why should I share? Examining social capital and knowledge contribution in electronic networks of practice. *MIS Quart.* 29(1):35–57.
- Wooldridge J (2002) *Econometric Analysis of Cross Section and Panel Data* (MIT Press, Cambridge, MA).
- Zhang X, Zhu F (2011) Group size and incentives to contribute: A natural experiment at Chinese Wikipedia. *Amer. Econom. Rev.* 101(4):1601–1615.
- Zhu F, Zhang XM (2010) Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *J. Marketing* 74(2):133–148.