"Popularity Effect" in User-Generated Contents: Evidence from Online Product Reviews

Abstract

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, and more important, what kind of reviews do such interactions induce? We study these questions using data from epinions.com, one of the largest product review websites where users can subscribe to one another. By applying both panel data and flexible matching methods, we find that as users become more popular, they produce more reviews and more *objective* reviews; however, their numeric ratings systematically change, and become more negative and more varied. Such tradeoff has not been previously documented, and has important implications for not just product review websites, but user-generated content sites as well.

Key Words: product reviews; user-generated content; online community; opinion leader; social media

1. Introduction

Online user reviews have become an increasingly important source of information for consumers. Many studies have shown that these reviews significantly affect consumer choices and therefore product sales (Aggarwal et al., Forthcoming, Berger et al., 2010, Chintagunta et al., 2010, Clemons et al., 2006, Dellarocas et al., 2007, Duan et al., 2008, Forman et al., 2008, Li and Hitt, 2008, Liu, 2006, Sun, 2012, Zhu and Zhang, 2010). However, it remains not well understood how reviews are generated in the first place. Several prior studies emphasized consumers' independent decisions; 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. While ratings have been shown to influence each other (Moe & Trusov, 2011), most online users are silent (Dellarocas and Wood, 2007) and only considered passive recipients of reviews. However, 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 an aggregate number that receives others' opinions, but are also individually visible to those who write. For instance, readers may "like" a review or share it on social media sites; they may rate them as helpful or unhelpful, or they can subscribe to select writers for their 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 empirical studies documenting the presence or absence of such effects. 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 interactions, i.e. subscription or "following." When a reader subscribes to a writer, contents generated by that writer will have priority over other writers when displayed to the reader. For review writers therefore, their subscribers (followers) create 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 tradeoff should be carefully weighed, and any induced "bias" should be recognized. For websites, they 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 Section

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. Section 6 contains post hoc analyses regarding linguistic features of product reviews. In section 7, we discuss the implications of this study, and some potential directional 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 dataset 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 to 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 "non-rating articles" in this paper[3]. What

[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 non-rating article could be a generic purchase guide (e.g. for buying a car).

makes the website particularly interesting is the "web-of-trust" (WOT) feature. Each user can choose to "trust" one or more other users, so that contents 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, reviews written by Jane will be displayed to John ahead of other reviews. This is highly comparable to "following" a user on Twitter or other social media. In addition, John may trust other users, and Jane may trust other users as well; this in turn creates a directed web-of-trust. 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 their 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 two important and expansive streams of research in the literature: *social influence* and *online word-of-mouth*. We cannot exhaustively survey all major studies in these fields. Rather, our goal is 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, Oh and Tan (2011) 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 disproportionally high level of influence on others' product choices. Various methods have been proposed to identify opinion leaders in a network (Iyengar, Van den Bulte and Valente, 2011, Trusov et al., 2010). Using data from epinions.com, Lu, Jerath and Singh (2011) provide insights into how ties are formed over time, and how opinion leaders emerge.

Our study can potentially fill a remaining gap in this burgeoning literature. While 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 contents interact with each other. Researchers in social psychology have identified that the mere presence of observers can change behaviors. Social psychologists refer to this 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, allowing the writers to easily "push" their writings to their followers. It seems natural, therefore, that the presence or absence of a captive audience can affect their behaviors. For product reviews, incoming ties may affect the writers' 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 consumer choices, product sales and even investor decisions (Aggarwal, Gopal, Gupta and Singh, Forthcoming, Clemons, Gao and Hitt, 2006, Dellarocas, 2003, Duan, Gu and Whinston, 2008, Forman, Ghose and Wiesenfeld, 2008, Gu et al., 2012, Liu, 2006, 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, Pavlou and Zhang, 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) find that the level of uncertainty surrounding transactions can have an influence on rating behaviors as well. Other scholars have examined how product positioning, such as niche versus hit products, affect the generation of reviews (Dellarocas et al., 2010). A common theme in this literature is that the decision to contribute reviews is a result of product characteristics that resulted in varying consumer experiences, or consumer characteristics.

A second category of studies focuses on how expressed product opinions may affect other product 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 by providing the median number of reviews contributed by community members, those who used to write less than the median write more. More recently, Wang, Zhang and Hann (2010) show that among users who are friends, the ratings provided by them influence each other as well. Our study takes a new perspective on the generation of product reviews. We study the effect of a larger online audience on a users' 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.

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

3.2 Hypotheses Development

We now develop the main hypotheses that we will test in this paper. While 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 due to its positive externality (Bolton et al., 2004, Chen, Harper, Konstan and Li, 2010). Being recognized as a useful or trustworthy source of information provides an important intrinsic motivation for writers of product reviews. 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. Therefore, receiving ties should encourage the writer to contribute more product reviews. While 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 such a status, so this effect is unlikely to apply to the majority of users[4]. 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. This nonlinearity can be captured in a quadratic term. We therefore hypothesize that:

H1: Receiving more incoming ties should increase the number of product reviews and non-review articles that a user will contribute 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 (Ofir and Simonson, 2001). Through lab experiments and field studies, Ofir and coauthors (2001) show that when consumers "expect to evaluate," they are likely to focus more on negative aspects of a product, resulting in more negative reviews. Hence, if H1 is supported, then the valence of reviews is also likely to decrease in the process as well. A second reason lies in the behavioral bias of readers toward negative reviews, and how writers of product reviews leverage such biases. It is widely observed that in many online communities, a small number of active users contribute the majority of contents (Kuk, 2006). For those active users, an important motivation is increasing popularity or obtaining higher status (Roberts et al., 2006).

[4] This is also the motivation that in our robustness tests we remove those with extremely high number of incoming ties, to ensure that these outliers or influential observations do not unduly bias our results.

Posting negative reviews serves this purpose, as readers tend to view negative opinions as being more useful, or smart (Amabile, 1983, Gibson and Oberlander, 2008, Moe and Trusov, 2011). Hence, if a writer is “strategic” enough, he or she will start with negative reviews. Receiving more incoming ties will only serve to confirm their prior believe, and they will have little incentive to deviate and post more positive ratings. They will either continue to be negative, or become even more negative. By contrast, for non-strategic writers, they may start with a range of opinions; some positive, some negative. With the same behavioral bias of readers, they gradually learn that negative opinions are more likely to attract followers, so they will become increasingly more likely to post negative reviews. Therefore, the overall effect of increasing the number of followers on a review writer should be negative: the more followers they have, the more likely that the writer will provide negative reviews. On the other hand, as the number of incoming ties increase, 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 therefore hypothesize that:

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.

Variance (measured as standard deviations) of ratings, on the other hand, has only recently been recognized as containing valuable information about products (Sun, 2012). We therefore also investigate whether and how the presence of an audience affects the variance of ratings generated by a review writer. However, since there has been little theoretical discussion on the antecedents to the variance of reviews, especially as defined from the perspective of review writers, we do not impost ex ante predictions on this, but leave it as an empirical question. We therefore hypothesize that:

H3: The variance of reviews that users generate will be affected by the number of incoming subscription ties that they receive.

In addition to these characterizations of the numeric ratings, we are also interested in how incoming ties affect the linguistic features of reviews, but leave this as an ex post analysis without positing specific hypotheses due to the lack of theories or published studies to draw on.

4. Data

We created automated agents to collect data on epinions.com in accordance with their robots.txt file. The first step of our data collection was to obtain the *universe* of users on its web-of-trust (WOT) network. To this end, we employed a "snowball" approach. Specifically, we started with the top 10 contributors in each main category on 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. We also obtained the date on which each tie was created[5], information about users on their profile pages, and information about the product reviews that

[5] Consistent with prior studies of epinions (e.g. Lu, Jerath and Singh 2011), we focus on ties created after 2002, as 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 been changed; for instance, many users in the WOT have been made anonymous.

they wrote (e.g., time stamp, length of reviews, numeric ratings, etc.). At the end, the dataset contains 92,094 user names (all users connected to the WOT), 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.

[Insert Tables 1 and 2 about here.]

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

For our empirical tests we employ panel data as well as matching methods to ensure robustness of our findings. In particular, the matching method not only allows us to estimate the treatment effect of having incoming ties, but also account for different levels of treatment *strength* nonparametrically, which we will describe later.

5. Empirical Analyses 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 non-rating articles provided by the user. For the main independent variable of interest, we focus on the user's "pure in-degree," or the number of non-reciprocated incoming ties. Although the proportion of reciprocated ties is small, those ties may be inherently different from the non-reciprocated 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. Number of non-reciprocated out-going 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.

A cross-sectional analysis of the data will lead to erroneous findings, as it does not account for potential endogeneity. For instance, while 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 (cf. Lu et al. 2011). We therefore turn to panel data models, and complement it with matching methods. Meanwhile, some unique features of our context and data assure the validity of these approaches. First, reverse causality and simultaneity bias is mitigated in the panel data, as 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 subscribe to its writer is quasi-random.

5.2 Dynamic Panel Data Model

We construct a panel dataset such that each unit of observation is a member, and each time period is one calendar month. The dataset therefore contains monthly observations about each user, including the number of subscribers or followers that the user had in each month, their activities such as number of product reviews that they wrote in that month, the mean and variance of these ratings, and so on. We first test for serial correlations using the method proposed by Wooldridge (2002) and implemented in Stata package XTSERIAL (Drukker, 2003). We find that for the number of ratings and number of non-rating articles, there is statistically significant first-order serial correlation, and it remains even if we use a binary indicator for writing any reviews (versus none at all) or non-review articles. Therefore, we turn to dynamic panel data models for these two outcome variables, and used the Arellano-Bond estimator (Arellano and Bond, 1991, Ghose, 2009) to estimate the effect of incoming ties on volume of reviews and non-review articles. This was 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[6]. These include the number of reviews, number of non-review articles, number of incoming ties, number of outgoing ties, and their overlaps[7]. As a robustness test, we also examined the binary outcome variable for generating any product ratings or non-ratings articles, respectively, in addition to the count measures. This specification is estimated using fixed-effect

---

[6] 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, as they do not provide meaningful variations in either the independent variable or dependent variable.

[7] 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 5 are from members that John also trusts. We hence separate the 5 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.

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 over-dispersion.

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. However, we note that 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 posted in that month by that user. We therefore estimate a Heckman selection model. 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[8].

5.3 Matching

To further test the robustness of the results obtained from the panel data setup, we take an entirely orthogonal estimation approach and consider matching instead (Heckman et al., 1998). The matching method helps overcome two potential 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 2 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" and it can be flexibly examined using the matching method.

[8] For example, we tested defining "selection" as writing at least 3 ratings so that the mean and standard deviation of ratings will only be calculated when there are at least 3 ratings from that user in that month.

Matching methods are increasingly popular among empirical researchers in IS. Broadly speaking, we first identify a user that received a certain treatment of interest. Here, a treatment is defined as receiving a number of incoming ties over a number of days (see next paragraph). We then identify another non-treated member that is highly similar to the treated. We repeat this process for all "treated" members, and then compare the outcomes of interest in these two groups of observations. We apply nearest neighbor matching using the Stata module NNMATCH (Abadie and Imbens, 2006), as has been applied in prior studies (Xue et al., 2011).

An added benefit of the matching method in our context is that we can estimate the treatment effect by looking at different levels of treatment *strength*, which has not been previously considered in the literature. More specifically, instead of defining "treatment" as receiving *any* incoming ties, we define it 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 3 incoming ties in 1 day, and B receives 3 incoming ties over 3 days, then the effect on behavior should be stronger for A than for B[9]. Once matched pairs are identified, we track their activities (contribution of ratings and non-rating articles, and the mean and standard deviation of the ratings) in the following 30 days[10]. 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.

---

[9] On the other hand, when X/Y ratio increases, the number of observed treatment units will decrease: All else equal, it is more likely for users to receive 3 ties over 3 days, than to receive 3 ties in just 1 day. This will reduce the number of available pairs for comparison.
[10] 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.

5.4 Discussion of Results

Our results from the dynamic panel data as well as the matching method are highly consistent. Tables 3-7 report the panel data results of various specifications for the four outcomes of interest, using different modeling techniques as appropriate. It should be noted that different columns may contain estimates from different empirical models and dependent variables[11]. Table 8 provides the results from the matching method.

[Insert Tables 3-8 about here.]

(1) **Volume of ratings and non-rating reviews**. Results from the panel data models provide support for H1: the coefficient on the number of incoming ties (pure in-degree) is positive and statistically significant, while 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 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 non-reciprocated 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 methods provide more straightforward interpretations. For instance, receiving 3 incoming ties in 3 consecutive days can, on average, motivate the user to provide 6.6 more new product ratings in the subsequent 30 days, and write 1 more non-rating article. The magnitude of

[11] 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.

this effect varies by the intensity of the treatment: if the number of incoming ties is spread out over 6 days instead of 3 (a weaker treatment), the increase is about 4. 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 an 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 to begin with. In other words, to encourage the sharing of consumption experiences, and if the emphasis is on the number of reviews or non-review articles, website administrators may be better off showcasing the reviews written by “up-and-coming” contributors rather than established “celebrities.” Given that writing product reviews is largely a public good (Chen, Harper, Konstan and Li, 2010), these results are particularly important for social media sites.

(2) **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, while 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. even 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 as more peers trust the user, he or she is indeed likely to write more negatively. The marginal effect is decreasing; however this effect is weaker due to larger standard errors of the panel data estimate. Easier interpretations come from the matching methods: Receiving 3 incoming ties in 2 days, on average, is followed by the user providing ratings about -0.03 stars lower. Since ratings range from 1 to 5 stars, one may argue that a decrease of 0.03 seems rather miniscule. Yet if a product is rated by a group of 30 writers that all

received 3 new incoming ties in 2 days, then the average rating on that product can be reduced by 1 star, which is by no means negligible. Alternatively, at the rate of 3 new ties every 2 days, assuming that this effect can scale to 60 days (multiplied by 30), the cumulative effect is also 1 fewer star out of 5. For website administrators, this difference may further magnify due to the increasing interactions among expressed opinions (Moe and Trusov, 2011). Therefore, when we look at the aggregate effect from multiple opinion writers and their cumulative effect over time, the impact can be substantial.

(3) **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; however, as the threshold increases, the number of usable observations (and therefore statistical power) decreases as well. From the matching results, we see that for the range of X-Y combinations that we estimate, 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 their 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, Pavlou and Zhang, 2006). Our results indicate that user interactions, such as the opportunity to subscribe to each other, may further encourage divergent views on products.

Before we conclude this section, we turn to the findings on some of the control variables in the panel data models. A user's number of outgoing ties (pure out-degrees), for instance, is fully controlled by the user themselves. 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 non-rating articles, suggesting that reciprocated ties – which may indicate deeper connections, acquaintances or friends – are also likely to motivate users to contribute. Its marginal effect is also decreasing. Lastly, holding the incoming and outgoing ties constant, the longer the user is on the site (time since registration), the fewer ratings and non-rating 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 their effort to increase their audience by acting negatively (Amabile, 1983), and differentiating these explanations can be an interesting area for future research.

6. Additional Analyses: Text Features of Product Reviews

Readers of product reviews don't just count the number of stars; they also read. Recent papers have shown the informational value of such textual features of product reviews (Archak et al., 2011, 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. We therefore explore the effect of audience on writing features in this section. We apply the same matching method to textual analyses by comparing the texts written by users who received the treatment (within the same 30-day window period that follows the treatment), to someone else that did not. There are many possible dimensions of texts that can be investigated; rather than trying to be comprehensive, our goal here is to explore some of the most well established characterizations: (1) positive sentiment; (2) negative sentiment; and (3) readability of reviews. Text sentiment 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) are less likely to be objective[12]. To obtain these two metrics, we use the LIWC package (Linguistic Inquiry and Word Count) (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). On the other hand, reviews should be easy to understand to sway consumer decisions. We therefore examine the readability of reviews using two popular 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). They are calculated using the following formulae:

$$\text{Lexical Density}^{13} = \left(\frac{Number\ of\ Unique\ Words}{Number\ of\ Words}\right) \times 100$$

[12] 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 2011 focus on movie ratings). We therefore indirectly assess the objectivity of texts by studying the portion of emotional words identified by LIWC.

[13] 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 19th, 2012).

$$\text{Gunning-Fog Index}^{14} = 0.4 \times \left(\frac{\mathit{Number\ of\ Words}}{\mathit{Number\ of\ Sentences}} + \frac{\mathit{Number\ of\ hard\ words}}{\mathit{Number\ of\ Words}} \times 100\right)$$

where "hard words" are defined as words with three syllables or more[15]. Lexical density (LD) measures the degree of information contained in texts. Higher density suggests that a text contains more information and more difficult to read. The Gunning-fog index, on the other hand, measures the estimated number of years of education that a reader needs to obtain to understand the text. Higher fog index suggests texts are more difficult to understand.

Once these calculations are done, we compare these linguistic metrics between treated and non-treated units using pair-wise t-tests. The magnitude of the difference, or effect size, can be seen from the mean difference, or a 95% confidence interval of the difference. Results from these tests are reported in Table 9.

[Insert Table 9 about here.]

Some interesting findings emerge from the results. 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 vs. 2.9% of all words for 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% vs. 1.99%) or negative (.51% vs. .83%). These suggest that users who are trusted by more peers tend to become more objective in their reviews. In other words, they increasingly sound like more of an authority rather than an amateur in their writings.

---

[14] Source: http://www.usingenglish.com/glossary/lexical-density-test.html, accessed July 19th, 2012.
[15] 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 19th, 2012).

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% vs. 40%), whereas the Gunning-fog index increases (9.5 vs. 9.3). However, the effect size on the Fog index is virtually non-existent, because the difference amounts to less than 3 months of education (12 months per year $\times$ (9.5-9.3)). Meanwhile, although one may argue that the decrease in lexical density suggest that reviews are becoming more readable, the difference is also modest, because LD scores below 40% are all considered low density (see the footnote on the previous page). Overall, these results suggest that readability of the reviews appear to be quite stable even when the writer gathers a large online audience. It is the words being used (emotional or non-emotional) that show a more interesting pattern as we previously discussed. But certainly, linguistic analysis is still an evolving field of research, and we may derive new insights as new tools and methods become available.

7. Implications and Future Research

A direct implication of our findings is that online user interactions, especially in the form of subscriptions that has become ubiquitous in social media, do affect user behavior. For websites that are trying to increase the sheer volume of activities on their sites (i.e. website traffic), our results confirm that user interactions do encourage more reviews or non-review articles to be generated. The decreasing marginal effect (H1) suggest that 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. Such efforts will also help induce a larger number of objective reviews on the website, which should arguably increase the reviews' informational contents. Meanwhile, these incoming ties also affect the valence and variance of reviews being generated. Marketers and website owners should carefully weigh these tradeoffs 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 participants answer independently, online product reviews reflect the interaction among expressed opinions (Moe & 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. 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 (Yuqing et al., 2012), especially how user interactions affect participation behaviors.

Our paper also contributes to the ever-growing literature on online word-of-mouth (Dellarocas, 2003, Forman, Ghose and Wiesenfeld, 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 online. Second, we are 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 extent studies that examine the textual content and linguistic features of user reviews (Lu, Jerath and Singh, 2011, Pavlou and Dimoka, 2006), measuring both the emotional content and readability of reviews.

More broadly, our results also address the "wisdom of the crowd" (Malone et al., 2010) phenomenon. User reviews are often cited as an example in the sense that even if one consumer's opinion is idiosyncratic, the overall estimate should be quite accurate when there are a sufficiently large number of such opinions. Recently, lab studies (Lorenz et al., 2011) show that social interactions and social influence can in fact 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 not allowed in lab experiments such as in Lorenz et al (2011).

The current study can be extended in several directions. For instance in our dataset, we do not observe the actual adoption or purchase behavior of the review-writers. It is possible that a 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 less negative. If we have information about their purchase behavior over time, we will be able to answer whether this change is due to (1) change in the products that they consume (i.e. selection); (2) same products, but the reviewer is becoming more stringent; or (3) same products and same opinions, but just modifying the “subset” of opinions to reveal online. This is an important question for further research, but beyond the scope of our current dataset.

It will also be interesting and worthwhile to extend our framework to other contexts of user-generated contents (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. While behavioral outcomes will differ from context to context, it is quite plausible that the larger audience will change the behavior of those being followed (e.g. risk preference for the experts’ stock choice on social investment platforms). These will be interesting future research questions.

Table 1: Summary Statistics of User Information (Cross-sectional, per user)

Variable	Mean	Minimum	Maximum	Standard Deviation	N
(a) Number of ratings	11.07	0	4094	51.36	92090
(b) Average rating given	3.98	1	5	.91	62344
(c) Standard deviation of ratings given	.79	0	2.83	.66	62344
(d) Number of non-rating articles written	1.26	0	438	7.61	92090
(e) In-degree (Trusted by)	6.60	0	2829	36.31	92090
(f) Out-degree (Trusting)	6.60	0	1830	27.56	92090
(g) Number of days since registration	3041.91	0	3667	699.49	92090

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)

	(a)	(b)	(c)	(d)	(e)	(f)	(g)
(a)	1.000						
(b)	-0.001	1.000					
(c)	0.111	-0.225	1.000				
(d)	0.553	0.013	0.093	1.000			
(e)	0.638	0.006	0.086	0.557	1.000		
(f)	0.458	0.023	0.103	0.448	0.615	1.000	
(g)	0.020	0.092	0.067	0.070	0.071	0.061	1.000

This table reports cross-correlation among the major variables. Full variable names of (a) through (g) can be found in Table 1 above.

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***	0.383***	0.532***
	(0.025)	(0.024)	(0.046)
Pure out-degree (natural log)	0.215***	0.207***	0.016
	(0.027)	(0.026)	(0.050)
Reciprocated ties (natural log)	0.775***	0.329***	0.438***
	(0.031)	(0.030)	(0.053)
Pure in-degree ^ 2	-0.374***	-0.236***	-0.028***
	(0.010)	(0.010)	(0.009)
Pure out-degree ^ 2	0.046***	-0.003	-0.044***
	(0.009)	(0.009)	(0.012)
Reciprocated ties ^ 2	-0.257***	-0.119***	-0.015
	(0.009)	(0.009)	(0.011)
Log number of ratings (t-1)	0.244***	0.195***	(N/A)
	(0.003)	(0.003)	(N/A)
Number of months on site (natural log)	-0.587***	-0.532***	-1.460***
	(0.009)	(0.008)	(0.026)
Intercept	2.701***	2.327***	1.361***
	(0.035)	(0.036)	(0.082)
N	138476	135153	161182

This table reports results on the volume of ratings using a panel data. The first column reports the results from 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 next table). 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 higher number of incoming ties is associated with 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. (* $p<0.1$, ** $p<0.05$, *** $p<0.01$)

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	0.007***	0.023***
	(0.000)	(0.001)
Pure out-degree	0.002***	0.012***
	(0.001)	(0.002)
Reciprocated ties	0.009***	0.008***
	(0.001)	(0.002)
Pure in-degree ^ 2	-0.000***	-0.000***
	(0.000)	(0.000)
Pure out-degree ^ 2	-0.000***	-0.000***
	(0.000)	(0.000)
Reciprocated ties ^ 2	-0.000***	-0.000***
	(0.000)	(0.000)
Number of months on site	-0.027***	-0.104***
	(0.001)	(0.002)
Intercept	-1.147***	1.660***
	(0.032)	(0.110)
N	121866	161182

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 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 higher number of incoming ties (pure in-degree) is associated with 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 1 rating or more	"Selection" defined as providing 2 ratings or more	"Selection" defined as providing 3 ratings or more
Pure in-degree	-0.079***	-0.112***	-0.112***
	(0.026)	(0.024)	(0.024)
Pure out-degree	0.024	0.032	0.032
	(0.023)	(0.023)	(0.023)
Reciprocated ties	0.042*	-0.030	-0.030
	(0.022)	(0.024)	(0.024)
Pure in-degree ^ 2	0.006*	0.009***	0.009***
	(0.003)	(0.003)	(0.003)
Pure out-degree ^ 2	0.007	0.008	0.008
	(0.005)	(0.005)	(0.005)
Reciprocated ties ^ 2	-0.017***	-0.002	-0.002
	(0.005)	(0.004)	(0.004)
Time on site	0.062**	0.053*	0.053*
	(0.031)	(0.029)	(0.029)
Intercept	4.348***	4.238***	4.247***
	(0.096)	(0.086)	(0.090)

This table reports the results of Heckman models 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 calculate if there are two or more ratings provided in that month; and the third, only calculate 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 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 2 or more ratings	"Selection" defined as providing 3 or more ratings
Pure in-degree	0.238**	0.209*
	(0.113)	(0.110)
Pure out-degree	-0.007	-0.039*
	(0.024)	(0.022)
Reciprocated ties	0.164**	0.162**
	(0.069)	(0.077)
Pure in-degree ^ 2	-0.018*	-0.017*
	(0.010)	(0.010)
Pure out-degree ^ 2	-0.019*	-0.010
	(0.010)	(0.009)
Reciprocated ties ^ 2	0.016	0.007
	(0.010)	(0.008)
Time on site	-0.447**	-0.355*
	(0.201)	(0.194)
Intercept	-0.463	-0.137
	(0.644)	(0.633)

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; and the second, only calculate 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)

Table 7: Panel Data Models for Incidence of non-Review Articles (log scale)

	Arellano-Bond Estimate for Number of Non-Rating Articles (log)	Arellano-Bond Estimate for Number of Non-Rating Articles (log, no outliers)	Fixed Effect Logit Model for 1(write any non-rating articles)
Pure in-degree	0.104***	0.019***	0.859***
	(0.006)	(0.005)	(0.168)
Pure out-degree	0.000	-0.013**	0.744***
	(0.007)	(0.006)	(0.191)
Reciprocated ties	0.111***	0.036***	2.574***
	(0.008)	(0.007)	(0.227)
Pure in-degree ^ 2	-0.077***	-0.016***	-0.083***
	(0.003)	(0.002)	(0.026)
Pure out-degree ^ 2	0.012***	0.014***	-0.160***
	(0.002)	(0.002)	(0.037)
Reciprocated ties ^ 2	-0.050***	-0.017***	-0.254***
	(0.002)	(0.002)	(0.037)
Number of non-rating articles, lagged	0.252***	0.125***	(N/A)
	(0.003)	(0.003)	(N/A)
Time on site	-0.037***	-0.019***	-1.979***
	(0.002)	(0.002)	(0.096)
Intercept	0.346***	0.104***	-5.612***
	(0.010)	(0.009)	(0.251)
N	138476	135153	161182

This table reports results on the number of non-rating 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 Arellano-Bond linear dynamic panel-data model, where the dependent variable is the log number of non-ratings 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 higher number of incoming ties is associated with higher probability of providing non-rating 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: Robustness Test: Effect of Receiving X Incoming Ties over Y days on Ratings and non-Rating Articles

X	Y	ΔΔ(Reviews Written)	ΔΔ (Mean of Ratings)	ΔΔ (Standard Deviation of Ratings)	ΔΔ (Non-review 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***

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 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 9: Ex post analysis: Effect of Receiving X incoming ties over Y days on the linguistic features of product 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

This table reports the effect of receiving 3 incoming ties in 2 consecutive days on the linguistic features of product reviews that users generate. Average treatment effect is identified as the expected difference between the treated group and the matched comparison group. All effects are statistically significant at 1% level, and we provide the standard errors, t-values (against null hypothesis of 0 difference), as well as the upper and lower bounds of 95% confidence interval for the treatment effect. Note that emotional words (positive and negative) as well Lexical Density are percentages, whereas the Gunning-fog index is interpreted as the number of years of education required for understanding the text.

References

A. Abadie, Imbens, G.W. 2006. Large sample properties of matching estimators for average treatment effects. *Econometrica*. **74**(1) 235-267.

J.G. Adair. 1984. The Hawthorne effect: A reconsideration of the methodological artifact. *Journal of Applied Psychology*. **69**(2) 334.

R. Aggarwal, Gopal, R., Gupta, A., Singh, H. Forthcoming. Putting Money Where the Mouths Are: The Relation Between Venture Financing and Electronic Word-of-Mouth. *Information Systems Research*.

T.M. Amabile. 1983. Brilliant but cruel: Perceptions of negative evaluators. *Journal of Experimental Social Psychology*. **19**(2) 146-156.

S. Aral, Muchnik, L., Sundararajan, A. 2009. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences*. **106**(51) 21544.

N. Archak, Ghose, A., Ipeirotis, P.G. 2011. Deriving the pricing power of product features by mining consumer reviews. *Management Science*. **57**(8) 1485-1509.

M. Arellano, Bond, S. 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies*. **58**(2) 277-297.

M.K. Bednar. 2012. Watchdog or lapdog? A behavioral view of the media as a corporate governance mechanism. *Academy of Management Journal*. **55**(1) 131-150.

J. Berger, Milkman, K.L. 2012. What makes online content viral?? *J. Marketing Res.* **49**(2) 192-205.

J. Berger, Sorensen, A.T., Rasmussen, S.J. 2010. Positive effects of negative publicity: When negative reviews increase sales. *Marketing Sci.* **29**(5) 815-827.

G.E. Bolton, Katok, E., Ockenfels, A. 2004. How effective are electronic reputation mechanisms? An experimental investigation. *Management Science*. **50**(11) 1587-1602.

J.M. Brett, Olekalns, M., Friedman, R., Goates, N., Anderson, C., Lisco, C.C. 2007. Sticks and stones: Language, face, and online dispute resolution. *Academy of Management Journal*. **50**(1) 85-99.

A. Cheema, Kaikati, A.M. 2010. The effect of need for uniqueness on word of mouth. *J. Marketing Res.* **47**(3) 553-563.

Y. Chen, Harper, F.M., Konstan, J., Li, X. 2010. Social comparisons and contributions to online communities: A field experiment on movielens. *The American Economic Review*. **100**(4) 1358-1398.

P.K. Chintagunta, 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.

E.K. Clemons, Gao, G.G., Hitt, L.M. 2006. When online reviews meet hyperdifferentiation: A study of the craft beer industry. *Journal of Management Information Systems*. **23**(2) 149-171.

C. Dellarocas. 2003. The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management Science*. **49**(10) 1407-1424.
C. Dellarocas, Gao, G.G., Narayan, R. 2010. Are consumers more likely to contribute online reviews for hit or niche products? *Journal of Management Information Systems*. **27**(2) 127-157.
C. Dellarocas, Wood, C.A. 2007. The Sound of Silence in Online Feedback: Estimating Trading Risks in the Presence of Reporting Bias. *Management Science*. **54**(3) 460-476.
C. Dellarocas, Zhang, X., Awad, N.F. 2007. Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive Marketing*. **21**(4) 23-45.
D.M. Drukker. 2003. Testing for serial correlation in linear panel-data models. *Stata Journal*. **3**(2) 168-177.
W. Duan, Gu, B., Whinston, A.B. 2008. Do online reviews matter? An empirical investigation of panel data. *Decision Support Systems*. **45**(4) 1007-1016.
C. Forman, Ghose, A., Wiesenfeld, B. 2008. Examining the Relationship Between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets. *Information Systems Research*. **19**(3) 291-313.
A. Ghose. 2009. Internet exchanges for used goods: an empirical analysis of trade patterns and adverse selection 1. *MIS Quarterly*. **33**(2) 263-292.
B. Gibson, Oberlander, E. 2008. Wanting to appear smart: Hypercriticism as an indirect impression management strategy. *Self and Identity*. **7**(4) 380-392.
B. Gu, Park, J., Konana, P. 2012. The Impact of External Word-of-Mouth Sources on Retailer Sales of High-Involvement Products. *Information Systems Research*. **23**(1) 182-196.
R. Gunning. 1969. The Fog Index after twenty years. *Journal of Business Communication*. **6**(2) 3-13.
J.J. Heckman, Ichimura, H., Todd, P. 1998. Matching as an Econometric Evaluation Estimator. *The Review of Economic Studies*. **65**(2) 261-294.
N. Hu, Pavlou, P.A., Zhang, J. 2006. Can online reviews reveal a product's true quality? Empirical findings and analytical modeling of Online word-of-mouth communication. *Proceedings of the 7th ACM conference on Electronic commerce* 324-330.
R. Iyengar, Van den Bulte, C., Valente, T.W. 2011. Opinion leadership and social contagion in new product diffusion. *Marketing Sci.* **30**(2) 195-212.
G.M. Kasper, Morris, A.H. 1988. The effect of presentation media on recipient performance in text-based information systems. *Journal of Management Information Systems*. **4**(4) 25-43.
J. Keegan, Kabanoff, B. 2008. Indirect industry- and subindustry-level managerial discretion measurement. *Organizational Research Methods*. **11**(4) 682-694.
G. Kuk. 2006. Strategic Interaction and Knowledge Sharing in the Kde Developer Mailing List. *Management Science*. **52**(7) 1031-1042.
X. Li, Hitt, L. 2008. Self selection and information role of online product reviews. *Information Systems Research*. **19**(4) 456-474.

Y. Liu. 2006. Word of mouth for movies: Its dynamics and impact on box office revenue. *J. Marketing*. **70**(3) 74-89.

J. Lorenz, Rauhut, H., Schweitzer, F., Helbing, D. 2011. How social influence can undermine the wisdom of the crowd effect. *Processings of the National Academy of Sciences*. **108**(22) 9020-9025.

Y. Lu, Jerath, K., Singh, P.V. 2011. *The Emergence of Opinion Leaders in Online Review Communities*. http://ssrn.com/abstract=1562245.

T.W. Malone, Laubacher, R., Dellarocas, C. 2010. The Collective Intelligence Genome. *MIT Sloan Management Review*. **51**(3) 21-31.

W.W. Moe, Trusov, M. 2011. The value of social dynamics in online product ratings forums. *J. Marketing Res.* **48**(3) 444-456.

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

P. Pavlou, Dimoka, A. 2006. The Nature and Role of Feedback Text Comments in Online Marketplaces: Implications for Trust Building, Price Premiums, and Seller Differentiation. *Information Systems Research*. **17**(4) 392-414.

J. Pennebaker, Chung, C.K., Ireland, M., Gonzales, A., Booth, R.J. 2006. *The LIWC Application*. Erlbaum.

J. Read. 2000. *Assessing vocabulary*. Cambridge University Press.

S.C. Rice. 2012. Reputation and Uncertainty in Online Markets: An Experimental Study. *Information Systems Research*. **23**(2) 436-452.

J.A. Roberts, Hann, I.-H., Slaughter, S.A. 2006. Understanding the Motivations, Participation, and Performance of Open Source Software Developers: A Longitudinal Study of the Apache Projects. *Management Science*. **52**(7) 984-999.

A.G. Sawyer, Laran, J., Xu, J. 2008. The readability of marketing journals: Are award-winning articles better written? *J. Marketing*. **72**(1) 108-117.

P.V. Singh, Tan, Y., Mookerjee, V. 2011. Network effects: The influence of structural capital on open source project success. *MIS Quarterly*. **35**(4).

M. Sun. 2012. How does variance of product ratings matter? *Management Science*. **58**(4) 696-707.

A. Susarla, Oh, J.-H., Tan, Y. 2011. Social Networks and the Diffusion of User-Generated Content: Evidence from YouTube. *Information Systems Research* isre.1100.0339.

D. Teichroew, Joslin, J.C., Trieb, S.E., Maisshall, W.S., Kriebel, C.A., Randall, D.L. 1967. Free for all. *Management Science*. **13**(6) B-378-B-382.

M. Trusov, Bodapati, A.V., Bucklin, R.E. 2010. Determining influential users in Internet social networks. *J. Marketing Res.* **47**(4) 643-658.

C. Wang, Zhang, M., Hann, I.-H. 2010. *Social influence in online product ratings*. St. Louis, MO.

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

M. Xue, Hitt, L.M., Chen, P. 2011. Determinants and outcomes of Internet banking adoption. *Management Science*. **57**(2) 291-307.

R. Yuqing, Harper, F.M., Drenner, S., Terveen, L., Kiesler, S., Riedl, J., Kraut, R.E. 2012. BUILDING MEMBER ATTACHMENT IN ONLINE COMMUNITIES: APPLYING THEORIES OF GROUP IDENTITY AND INTERPERSONAL BONDS. *MIS Quarterly*. **36**(3) 841-864.

F. Zhu, Zhang, X.M. 2010. Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *J. Marketing*. **74(**2) 133-148.