THE ROLE OF SOCIAL NETWORKS IN ONLINE REVIEWING

Zhihong Ke, Gatton College of Business and Economics, University of Kentucky, Lexington, KY, USA, zhihong.ke@uky.edu
De Liu, Gatton College of Business and Economics, University of Kentucky, Lexington, KY, USA, de.liu@uky.edu

Abstract

Online reviews are a dominant resource for consumer decisions but what leads users to write reviews remains largely unexamined. Extant research on user content generation has primarily focused on what motivates users to contribute content, and less on the effects of informational and social environment surrounding these users. The aim of this study is to examine how a user’s contribution to an online review platform is affected by reviews of his/her friends from both informational and social perspectives. We expect that information, reciprocity, and social comparison are primary drivers for contribution. Among friends who wrote reviews, we predict that those who carry redundant information have less effect on the focal user’s contribution, whereas those who are strong-tie friends of the focal user have a stronger effect. Furthermore, we expect that users’ status moderates these effects such that an elite user responds more positively to friends who carry redundant information, and more negatively to those who are strong-tie friends, compared to non-elite users. Our expected findings hold implications for online review platforms in terms of highlighting the most relevant reviews generated by one’s peers.

Key words: Online reviews, Social Influence, User-Generated Content
1. INTRODUCTION

Online reviews of products, services, and businesses have become a dominant source of information for consumers. According to ComScore (2007), nearly one out of every four U.S. Internet users (24 percent) reported using online reviews prior to paying for a service delivered offline. More than three-quarters of online review readers reported that the reviews had a significant influence on their purchases. A 2013 report by Dimensional Research concluded that approximately two-thirds of U.S. consumers reported reading online reviews, among them, an overwhelming 88 percent said that their buying decisions are influenced by online reviews (Dimensional Research 2013). Numerous prior research has established the link between online reviews, also known as digital Word-of-Mouth, and sale performance (e.g., Chevalier and Mayzlin 2006, Cheung, Luo et al. 2009, Lu, Ba et al. 2013). On the other hand, few have examined the design of such online review systems (e.g., Dellarocas, Gao et al. 2010, Lu, Jerath et al. 2011). With the exception of few studies (e.g., Wang 2010, Goes, Lin et al. 2014), what leads online users to write reviews is largely unexamined.

Despite the ubiquity of online review systems, online reviews, especially creditable ones, are still a scarce resource. According to Maritz Research (2013), the ratio of U.S. consumers who read to those who contributed reviews is around 4 to 1. Major online review platforms have taken vastly different approaches to cultivating online reviews. For example, Epinions employs a revenue sharing strategy to elicit online reviews. Similarly, Amazon offers financial rewards to those top reviewers with the same aim. In contrast, CitySearch recognizes key contributors as experts to encourage their participation in review generation. Similarly, Yelp does not pay users to write reviewers but adopts a community strategy where users are encouraged to make friends with one another, vote on other reviews, or send compliments to other users. The success of unpaid review communities suggests that social interactions amongst reviewers may play an important role in review generation. The goal of this paper is to investigate various ways social interactions between reviewers may influence review generation in unpaid reviewing communities.

We view users of online review platforms as both consumers and producers of information (i.e., reviews). Explicitly recognizing these two different roles can lead to a more holistic view of the environment these users are situated in and thus offer us a chance to test competing hypotheses about the factors that influence their review generation. In this paper, we primarily focus on the quantity, or the number of reviews written. We believe that understanding the number of reviews produced is an important first step and prerequisite for other important questions such as the quality of reviews.

While there are few studies of review writing, the user-generated content (UGC) literature has extensively examined the incentives of voluntary contributors in other contexts such as open source software (e.g., Lerner and Tirole 2002, Franke and Hippel 2003), crowdsourcing (e.g., Huberman, Romero et al. 2009, Zhang and Zhu 2011), Question/Answer platform (e.g., Chen, HO et al. 2010, Tausczik and Pennebaker 2012, Jabr, Mookerjee et al. 2014), and content-sharing (e.g., Xia, Huang et al. 2007, Rui and Whinston 2011). Primarily from the perspective of content producers, previous research has shown that users are motivated by intrinsic or extrinsic motives such as the need for attention, recognition, and reputation (e.g., Nam, Ackerman et al. 2009, Rui and Whinston 2011, Zhang and Zhu 2011), they gain from content consumers, and by other-concern such as giving and helping content consumers (e.g., Cornes and Sandler 1984, Steinberg 1987, Price, Feick et al. 1995). In contrast, a separate literature on information diffusion has primarily treated users as consumers of information, therefore characteristics of information sources and information channels are studied and demonstrated (e.g., Watts and Dodds 2007, Susarla, Oh et al. 2012). However, users of online review platforms (and indeed in many other peer-to-peer communities) are both consumers and producers of information, thus it is necessary to combine and contrast the two perspectives. We argue that the information producer and consumer perspectives suggest different sets of environmental factors for reviewing behavior. We attempt to distinguish these factors, thereby shed light on the underlying mechanisms for review contribution.

Different from the studies of user motivations, we focus on the information and social environment surrounding users of online review platforms (ORPs), and examine the roles of different
information sources and the factors that moderate their importance. This perspective lends itself well to the design of ORPs. Modern ORPs surround their users with information contributed by other users (e.g., live feed of friend activities and suggested reviews for nearby stores). It is important to understand what type of information matters and how social relations moderate the value of information so that ORPs can design a more productive environment for reviewers.

We are also distinct from prior studies of social networks because, rather than modeling social relations as a first-order source of influence, we focus on the information flow between social actors and model social relations as moderators of the value of information. We believe actual information flow, which are now becoming widely accessible to researchers, are carriers of social influence and thus should be the focus of social influence studies.

We address these issues using a dataset on Yelp (www.yelp.com) users. Yelp is a rapidly growing ORP for local businesses. Yelp supports its legion of unpaid users with extensive social network features, including support for friends, following (as fans), compliments, online chatting, and voting on reviews written by peers. Yelp additionally recognizes a group of “elite” users each for their extraordinary contribution and invites them to local community events. Yelp provides information-rich user interfaces for both web and mobile users – users can view a live feed of information on “near-by” businesses and friends’ recent activities. These distinct design features make Yelp a rich context for studying the effect of social influence on online reviewers.

Our work makes several contributions to existing literature. We combine a content producer’s perspective and a content consumer’s perspective with an aim to develop a holistic understanding. This provides an opportunity to contrast these two perspectives and see how they may work together in ORPs. Second, our work examines various factors at the level of information flow between users, thus complements and extends existing literature that focuses on higher-level motivations and social network characteristics. Finally, we further clarify the mechanism of social influence by testing a model where social relations serve as moderators of information values.

2. RELATED WORK

Our work is closely related to a rich body of literature on user-generated content (UGC). Most prior empirical studies of UGC have concentrated on the relationship between content generation and its appreciation in the forms of attention, recognition, and reputation (e.g., Nam, Ackerman et al. 2009, Rui and Whinston 2011, Zhang and Zhu 2011), and on examining content producers’ desire to help consumers (e.g., Cornes and Sandler 1984, Steinberg 1987, Price, Feick et al. 1995). Less attention has been given to the social relationship amongst peer users/contributors, which is the main focus of our work.

Our study also has a close connection to literature examining the role of social influence in social networking sites such as YouTube and Epinions. Existing research examining the effect of social influence on content generation have observed two contrasting effects: reinforcement effect and substitution effect. For example, Xia, Huang et al. (2007) showed that sharers behave in reciprocal manner: the greater benefits they get from the network, the more likely they continue contributing (i.e., reinforcement effect), whereas Asvanund, Clay et al. (2004) studied participants’ behaviors across multiple file sharing networks, and found evidence of free riding (i.e., substitution effect). Previous findings on different directions of social influence inspire our study. Unlike previous studies that use aggregate measures of social influence, we take measures on each pair of relationship, which allows us to better clarify the nature and mechanisms of social influence in ORPs.

Finally, there is rich literature on using social networks to predict behaviors. It focused either on the relational aspect of social relations, for example, Nair, Manchanda et al. (2010) documented that physician prescription behavior is significantly influenced by the behavior of research-active specialists, or “opinion leaders,” in the physician's reference group; or it may focus on the informational aspect of social relations. For example, Susarla, Oh et al. (2012) showed that informational channels have a significant impact on the rate of diffusion of user-produced video
content on YouTube. Our work enriches this stream of research in two ways: we combine informational and relational arguments; we also study a new context, online review networks, rather than consumer networks.

3. THEORETICAL BACKGROUND AND HYPOTHESES

DEVELOPMENT

3.1 The Role of Contributing Friends

Being aware of a store is a prerequisite for visiting and eventually writing a review of it. One way a user of an ORP is affected by her contributing friends is through awareness. In ORPs, reading reviews by others is an important way for users to discover stores. Given a large number of reviews posted daily on stores around them, users are more likely to pay attention to reviews by friends, whom they trust and share interests with. ORPs often provide formal mechanisms for users to be alerted by friend reviews. Reading friend reviews thus increases one’s chance of visiting the store and writing one’s own review.

Friends’ review contributions may alternatively affect a user’s contribution through reciprocity. Researchers (e.g., Connolly and Thorn 1990, Xia, Huang et al. 2007) have used the concept reciprocity to explain voluntary contribution. When people expect to receive benefit from another, they tend to have the incentive to offer benefit as well. Furthermore, Emerson (1976), Ekeh (1974), and Yamagishi and Cook (1993) noted that when the other party is a group instead of an individual, generalized reciprocity applies, which suggests that people contribute because of the “obligation” to reciprocate with members in a group. In other words, people contribute reviews because they are grateful for the reviews they have enjoyed from their friends and feel obligated to pay back by writing their own reviews.

While the previous two explanations are primarily from the perspective of content consumers, a content-producer perspective may also offer an explanation. As a content producer, a user may compare her performance with those of her friends. Put differently, when a user observes that her friends are posting, social conformity will drive her to post also to keep up. Deviating from the social norm may adversely affect one’s reputation among her circle of friends. All the three mechanisms, awareness, reciprocity, and conformity would predict a positive effect of contributing friends:

Hypothesis 1: A user’s contribution in period t is positively related to the number of contributing friends in period t-1.

3.2 The Role of Non-Contributing Friends

When a friend has not posted new reviews, a user does not receive any new information and thus awareness and reciprocity-based arguments would predict that a non-contributing friend has no effect on a user’s review contribution behavior. The conformity argument, however, would suggest a negative effect: when one’s friends do not write reviews, a user perceives a lower norm for review contribution, thus the user would reduce her own contribution to match the perceived norm (of not contributing). Hence, conformity alone predicts:

Hypothesis 2: A user’s contribution in period t is negatively related to the number of non-contributing friends in period t-1.

3.3 Redundancy

From an information consumer’s perspective, a user values reviews that are novel. When a user reads a review about a new or different kind of store, she is more likely to check it out and follow with a review. By this argument, a friend who reviews similar stores as the user is less likely to trigger review contributions from the user, than a friend who reviews dissimilar stores. Moreover, a friend
review of the same kind of stores that the user reviews is less likely adds additional value to the user and thus less likely triggers reciprocation.

From a producer’s point of view, people tend to share their experiences that make them look intelligent and unique (Dichter 1966). For example, Ling, Beenen et al. (2005) found that people in an online movie recommender community will contribute more when receiving personalized information that their contributions would be unique. Similarly, feedback on the uniqueness of users’ contributions to a movie-rating website stimulated participation (Ludford, Cosley et al. 2004).

Thus the awareness, reciprocity, and uniqueness arguments predict:

**Hypothesis 3**: A user’s contribution in period t is less affected by the number of contributing friends in period t-1 who review similar stores, than by the number of contributing friends who review dissimilar stores.

### 3.4 Strength of Ties

Strength of tie refers to a combination of amount of time, emotional intensity, intimacy (mutual confiding), and reciprocal services that characterize the tie (Granovetter 1973). Weak ties are infrequent and distant (Hansen 1999), whereas strong ties are frequent, long-lasting and affect-laden (Krackhardt 1992). Drawing from the social network literature, we present two countervailing arguments for the effect of strong ties on a user’s review contribution. On one hand, strong tie may be associated with redundant information (Granovetter 1973), for strong tie friends likely review similar stores and have similar tastes. By the information redundancy argument, reviews contributed by strong-tie friends are less likely to trigger subsequent reviews. On the other hand, strong ties also coexist with stronger social identification. Social identity is the individual’s self-concept derived from perceived membership of social groups (Hogg and Vaughan 2005). Social identity theory (Tajfel 2010) suggests that content generators are driven by a desire to establish a distinct identity via first self-categorization (Dellarocas, Gao et al. 2010), then self-expression (Schau and Gilly 2003), self-presentation (Miller and Edwards 2007), and self-performance (Liu 2007, Van House 2007) in order to defend or pursue the distinct identity in their self-categorized groups. A strong identification with each other strengthens the reciprocity. When a member of a strong-tie group contributes, the rest of the members more likely respond in kind to shore up their support for common causes. Therefore, the social identity argument would predict a positive reinforcement effect by strong-tie friends.

In cases of costly actions, the reinforcement effects of strong ties have been shown to dominate the information redundancy effect. For example, strong ties have been shown to better facilitate knowledge transfer than weak ties (Hansen 1999). Online peer-to-peer lenders more likely follow strong-tie friends to lend money to a third party (Liu et al 2014). Writing a review requires considerable effort from a user, thus we hypothesize that strong-tie friends have stronger effect than weak-tie friends while acknowledging that countervailing forces exist.

**Hypothesis 4**: A user’s contribution in period t is more affected by the number of contributing friends in period t-1 who have stronger ties with the focal user, than weaker ties.

### 3.5 Status

So far we have treated users of ORPs as homogenous entities. In reality, users at different level of involvement with ORPs may behave quite differently. An entry level user’s opinions are less trusted, thus, as a content producer, she may add very little value by writing redundant reviews. An elite user’ opinions are trusted and valued by many users. Even though there may already be reviews by other users, an elite user’s opinion will still be desirable as an authoritative voice. Thus, we expect an elite to be more likely to write reviews when receiving similar reviews by her friends.

**Hypothesis 5**: A user’s elite status dampens the negative effect of redundancy, that is, an elite user is more positively affected by her contributing friends who review similar stores than non-elite users are.
Earlier we have argued that a strong-tie relationship can overcome the information redundancy and enable a reinforcement effect between friends. In the case of elite users, such a social-identity based argument is likely weakened. An elite user has already established herself as a leader of ORPs, thus reciprocating other members as a way of becoming accepted into the social group is less of a concern. From producers’ perspective, strong-ties of an elite user are more likely leading producers of reviews. An elite user are more likely to differentiate herself from such peers (Snyder 1980). Hence, a producer’s perspective would predict:

**Hypothesis 6**: A user’s elite status dampens the positive effect of strong ties, that is, an elite user is less positively affected by her strong-tie contributing friends than non-elite users are.

### 4. DATA

#### 4.1 Data Collection

We test our hypotheses using Seattle dataset. This is because according to Wang (2010), the number of restaurants (2,714), the total number of reviews (66,465), and the number of review per restaurants (24.5) in Seattle are close to the mean of 21 popular cities listed on front page of Yelp during that data collection period.

Restaurants are the most popular one among 22 categories on Yelp across all cities. Our data from searching “Seattle” and “restaurant” cover the vast majority of reviewers in Seattle, which included 7,845 non-elite users and 551 “elite” users. “Elite” users are labeled as such by Yelp based on the quantity of reviews as well as other criteria. They produce the majority of reviews on Yelp.

#### 4.2 Data Description

Our full dataset contains the complete information on 28,566 friends of 551 elite reviewers in Seattle between February 2013 and July 2013, including the compliments received by the user (sender, time, and type), the user’s friends, reviews (time and content), the number of votes, the number of fans, and other information. Among the 28,566 users, we removed those who did not contribute any reviews for restaurants in the Washington state, and those who were inactive (i.e., did not contribute any reviews in 3 consecutive months). 669 reviewers and their active friends (1,277 reviewers), who have written reviews for restaurants in Washington State, remained for analysis. These 669 reviewers and their friends (1,277 reviewers) generate 10,326 dyadic pairs, and produced a total number of 15,732 reviews for 7,221 restaurants in Washington State. About 660 reviewers at each period (a week) remained and a total of 20 periods generated 13,258 observations. Table 1 provides a summary statistic of user characteristics.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Definition</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td># of Compliments</td>
<td>The total number of compliments received</td>
<td>0</td>
<td>56,447</td>
<td>294.45</td>
<td>2310.93</td>
</tr>
<tr>
<td># of Fans</td>
<td>The total number of fans followed</td>
<td>0</td>
<td>507</td>
<td>8.00</td>
<td>25.25</td>
</tr>
<tr>
<td># of Reviews</td>
<td>The total number of review written</td>
<td>0</td>
<td>1,573</td>
<td>175.01</td>
<td>189.09</td>
</tr>
<tr>
<td># of Friends</td>
<td>The total number of friends had</td>
<td>0</td>
<td>4,325</td>
<td>88.78</td>
<td>243.39</td>
</tr>
<tr>
<td># of Votes</td>
<td>The summation of cool, funny, and useful votes</td>
<td>0</td>
<td>36,970</td>
<td>882.90</td>
<td>2,262.14</td>
</tr>
<tr>
<td>Yelp Months</td>
<td>The number of months residing in Yelp</td>
<td>2</td>
<td>97</td>
<td>44.12</td>
<td>19.49</td>
</tr>
<tr>
<td>Average # of Reviews Per Week</td>
<td>The average number of reviews per month written</td>
<td>0</td>
<td>33</td>
<td>0.59</td>
<td>1.51</td>
</tr>
</tbody>
</table>
5. **EMPIRICAL MODEL**

5.1 **Model Selection**

A panel-data logit and a panel data negative binomial model (NB) will be appropriate to model the social influence in question, as the dependent variable is either binary (whether a reviewer posted new reviews) or a count variable (the number of reviews written).

5.2 **Independent and Control variables**

**Independent Variables**

For Hypothesis 1 and 2, we will use the number of contributing and non-contributing friends in the last period, controlling for the total number of reviews post by contributing friends, as a measure of the role of contributing and non-contributing friends. This measure allows us to examine the effect of review contribution norm (i.e., reciprocity and conformity) in a reviewer’s network.

For Hypothesis 3 and 4, we will use two relationship attributes: redundant information and strength of tie. Redundant information is approximated by the number of reviews the pair wrote for similar businesses and stores in terms of store popularity, perceived quality, and price range. For example, if reviewer A and reviewer B writes many reviews for the similar stores in terms of store popularity, perceived quality, or price range, they are considered providing redundant information to each other. To calculate these similarities, we first establish a user’s profile in term of store popularity, perceived store quality, and price range. We categorize all stores in our dataset into five popularity levels based on the total number of reviews they received. For example, if a store is reviewed by more than 80 people, it is classified as a level-5 store. We count the number of stores reviewed by each user at each popularity level and store the five numbers in a vector as the reviewer’s store popularity profile. Store perceived quality is captured by the average ratings received. Yelp offers reviewers a 5-point scale to rate the stores. We similarly calculate counts of stores for each user at each quality level and construct a store quality profile for each user. The focal stores are also categorized into four price ranges by Yelp. We similarly construct a store price profile for each user using a 4-item vector. With user profiles on store popularity, quality, and price range, we proceed to calculate similarity between any pair of users along the three dimensions. Given the vector-based profiles, we use cosine similarity, one of the most widely adopted similarity measure. The cosine similarity between i’s and j’s in period t along a particular dimension is calculated as follows:

\[
SIM_{ijt} = \frac{\sum_{k=1}^{N} (R_{ikt} \cdot R_{jtk})}{\sqrt{\sum_{k=1}^{N} (R_{ikt})^2 \sum_{k=1}^{N} (R_{jtk})^2}}
\]

where

\( R_{ikt} = \) the number of reviews reviewer i in period t writes for level k stores (in terms of population, perceived quality, or price range);

\( R_{jtk} = \) the number of reviews reviewer j in period t writes for level k stores (in terms of population, perceived quality, or price range); and

\( N = \) number of dimensions the focal vector has.

After computing the similarity scores for store popularity, perceived quality, and price range respectively, we average the 3 similarity scores to obtain a single similarity index. This similarity...
index is used to classify a pair as having similar store preferences (i.e. redundant information carriers to each other) if the index value exceeds one standard deviation above the mean.

Strength of tie is quantified by social interaction between a pair. The social interaction is measured by the number of compliments between the pair. Yelpers can send compliments to their friends and peers over their reviews, photos, and tips. A compliment must be accepted by a recipient to be shown in the recipient’s profile. Mutual compliments are strong indicators of mutual affection and meaningful interactions between a pair of users. The average compliments each user received in our dataset are 294 with a range from 0 to 56,447. We use Jaccard Index to measure the number of compliments between a pair. The Jaccard Index between reviewer i and j in period t is calculated as follows:

\[ J_{ijt} = \frac{|x_{it} \cap x_{jt}|}{|x_{it}|} \]  

where

- \( |x_{it} \cap x_{jt}| \) = the number of compliments reviewer i and j in period t send to or receive from each other;
- \( |x_{it}| \) = total number of compliments reviewer i in period t send or receive.

Accordingly, we consider a pair to have a strong tie if the index value exceeds one standard deviation above the mean.

In each period, we group each user’s friends based on whether they are redundant-information and strong-tied friends and whether the friends have contributed in the last period. The number of cases are calculated for all users and for elite users only. As Table 2 shows, about half of the cases (6,151 elite observations out of 13,258 observations) are elite users in our dataset, and these elite users have more friends in each group. For example, the average number of friends who wrote reviews for all users is 6.38, whereas that for elite users is 11.08.

<table>
<thead>
<tr>
<th>Number of friends by groups</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Observation</th>
</tr>
</thead>
<tbody>
<tr>
<td>All users</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Review-contributing friends</td>
<td>0</td>
<td>289</td>
<td>6.38</td>
<td>13.01</td>
<td>13,258</td>
</tr>
<tr>
<td>- Strong-tied &amp; contributing friends</td>
<td>0</td>
<td>75</td>
<td>2.09</td>
<td>4.16</td>
<td>13,258</td>
</tr>
<tr>
<td>- Redundant-info &amp; contributing friends</td>
<td>0</td>
<td>56</td>
<td>1.57</td>
<td>3.49</td>
<td>13,258</td>
</tr>
<tr>
<td>Elite users</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Review-contributing friends</td>
<td>0</td>
<td>289</td>
<td>11.08</td>
<td>17.54</td>
<td>6,151</td>
</tr>
<tr>
<td>- Strong-tied &amp; contributing friends</td>
<td>0</td>
<td>75</td>
<td>3.65</td>
<td>5.58</td>
<td>6,151</td>
</tr>
<tr>
<td>- Redundant-info &amp; contributing friends</td>
<td>0</td>
<td>56</td>
<td>2.58</td>
<td>4.7</td>
<td>6,151</td>
</tr>
</tbody>
</table>

Table 2: Descriptive Statistics of Dyad Relationship Groups

Control Variables

In our model, we will control the content generation tendency of a reviewer, which is captured by the last period generation, the recent production, the aggregate generation, and the latest inactive days. To control for the appreciation from content consumers, we calculate the aggregate number of fans, votes, and compliments, and the number of incremental compliments the reviewer received by the end of last period. The experience of a reviewer is measured by the number of months that the reviewer had been on Yelp by the end of last period.

Our data analyses will be based on the above measures. To avoid endogeneity concerns due to reverse causality, we will use last period variables as repressors for the contribution in the current
period. To alleviate omitted variable bias caused by unobserved individual differences, we test a fixed-effect model. We also include time dummies to control for temporal shocks that affect all users. Our analyses are ongoing and we hope to report our results at the time of the conference.

6. CONCLUSIONS

We explore various ways social interactions between users may influence review generation in unpaid reviewing communities. Specially, we, from both informational and social perspectives, examine how a user’s contribution to an online review platform is affected by reviews she received from her friends. By combining and contrasting the content producer’s perspective and content consumer’s perspective, examining the role of information environment, and testing that information flow as the conveyor of social influence, we hope to bring a broad set of implications for the design of online review platforms.

References


Nam, K. K., et al. (2009). Questions in, knowledge in?: a study of naver's question answering community, ACM.