The Influence of Social Capital in an Online Community on Online Review Quality in China

Qiuju Li
University of International Business and Economics, Beijing, China
20131450879@uibe.edu.cn

Jinhong Cui
University of International Business and Economics, Beijing, China
cuijinhongwx@hotmail.com

Yun Gao
University of International Business and Economics, Beijing, China
20131450905@uibe.edu.cn

Abstract

Collaboration is to make full use of the advantages of each one, not only saving time but also saving resources. As a form of collaboration, online reviews attract attentions of scholars and businessmen in China. However, previous research mostly explores how review quality affects product sales or how review content influences quality of online review. In this study, we present how social capital of reviewers affects the quality of review. Based on the social capital theory and social network theory, we consider a reviewer’s social network, indegree and outdegree, experience, and activity, and build a theoretical model for review quality. We use 1764 reviews in an online community in China to test our hypotheses through the Tobit regression method. The research results not only make theoretical contributions to research on online review quality, but also help enterprises manage and apply online reviews effectively.

Keywords: Social Capital  Reviews Quality  Content Similarity  Social Network

1. Introduction

With the development of technology, it is getting easier for people to contact each other. Virtual teams become possible, and collaborative projects become more and more popular. The existing form of such collaboration includes open source software (OSS) development, crowdsourcing, Wikipedia, etc. Online product reviews is a kind of collaboration in social media environments, and reviewers form a “virtual team” whose aim is to let users know products well. Online review is an antecedent of online collaboration. Online communities/virtual teams and collaboration are also strongly associated [1] [2]. Review quality is important to online communities. In 2009, a survey made by Forrester Research Inc. showed that 91% of the respondents would consider online reviews, blogs, or other user generated information before they buy a new product or service, and 46% of people will be affected by such reviews before purchasing. Since reviews have great influence on purchasing decisions, the quality of reviews becomes a concern. It can not only help users find helpful reviews quickly, save time and cognitive cost [3] [4] [5], but also show the extent of reviews’ recognition and influence [6] [7]. There is a growing body of papers and studies related to user generated online reviews, mainly focusing on three aspects, including the influence of online reviews on the sales of goods, the helpfulness of online reviews, and the propagation mechanism of online reviews [8] [9] [10]. Meanwhile, with the huge amount of online reviews, users are more concerned with finding what are the helpful reviews.

Many scholars have studied the helpfulness of online reviews. Most of them consider factors that influence the quality of online reviews from three aspects: content of reviews, star ratings, and reviewers’ experience. Some of researches focus on reviews’ content, including length and readability of reviews [11] [12]. Some others use text mining methods to find out special words that could influence the helpfulness of reviews [3] [13]. Papers focusing on star ratings consider variances of ratings or ratings extremity [11] [12]. Studies with a focus on reviewers’ experience consider rank of reviewers or reviewers’ popularity [14] [15]. Mudambi et al. [12] pointed out that reviews with extreme ratings are less helpful than reviews with moderate ratings and review depth has a greater positive effect on the helpfulness of the review for search goods than for experience goods. Recent research uses data mining methods to analyze how special words, such as emotional expression, characteristics of products, affect reviews’ helpfulness. Yin [13] discovered that emotional expression of anger or anxiety in online reviews influence their helpfulness. Cao et al. [3] found that semantic characteristics of
reviews are influential in affecting how many helpfulness votes reviews receive.

Although some scholars have studied helpfulness from a view of reviewers’ experience, there is few research studied from a reviewers’ social capital perspective. This paper tries to find factors that influence reviews’ quality from the view of reviewers’ social network characteristics and build a model for reviews’ quality based on the social capital perspective. We draw on social network theory and social capital theory to explore how reviewers’ popularity, the groups’ characteristics in which a reviewer takes part, reviewers’ experience, and reviewers’ activity affect reviews’ quality.

We have organized the rest of this paper as follows. First, we present the theoretical background, then we provide hypotheses and a conceptual model. After describing our data source, we explain our method and present results of our data analysis. Finally, we conclude with a discussion on the implications for theory and practice.

2. Theoretical Background

2.1. Social Network Theory

A social network is a relatively stable relationship system in which every member interacts with others frequently. In 1954, Barnes [16] used social network for the first time in his book “Human Relations”. A social network consists of one or more finite actor sets and one or more relationships between them. Studies focusing on social networks prefer the relationship between actors to actors’ attributes. Previous research can be divided into two types. One type pays more attention on an individual’s connections and position in a network [17], such as bonding ties, bridging ties [18], centralization, structural holes [19]. The other type concentrates on the whole network as a unit and delves into the density of the network, cohesive subgroup, structural equivalence, etc [20]. The relationships that have been studied vary from friends, colleagues, to families, etc. Social network analysis uses relationships to describe the structure of a group, and then explores the impact of this structure on the group’s operations and its impact on individuals within the group.

There is a special relationship between members who take part in the same group in online community. It is a kind of bridging ties, members in the same group come from different social networks and information they spread is different. By joining groups reviewers can form social networks, and social networks will affect individuals who are in it. This study argues that there should be influence of reviewers’ social relationships on review quality through the characteristics of groups that a reviewer participates.

2.2. Social Capital Theory

Social capital refers to resources embedded in a social structure that are accessed and/or mobilized in a purposed action [21]. It was brought into the field of sociology by French sociologist Bourdieu in 1977. Nahapiet et al. [22] divided social capital into three dimensions: structure dimension, relational dimension and cognitive dimension. They defined structure dimension as the overall structural model of social connections among people, which can be measured by centralization, including indegree, outdegree, closeness, betweenness, information centrality, etc.

There are few studies on the impact of social capital on review quality, but some scholars have illustrated the relationship between user generated content (UGC) and social capital. From the start, past research have considered how social capital affects the quantity or quality of UGC from three dimensions qualitatively [23] [24] [25]. Most of them use Likert scale, and the factors they found are general and not practical. Later, scholars try to explore detailed factors from each dimension (structure dimension, relational dimension, and cognitive dimension). Current research concentrates on the structure dimension because it is easy to be measured. What current research studied include open source software [26], collaboration UGC [27], YouTube video sharing, etc [28]. In past research, it was concluded that the higher centrality degree, the higher quality of UGC [29]. Scholars also found an “inverted-U” shape relationship between centralization and the quality of UGC, meaning that if the centrality degree is too high, it will have negative influence on the quality of UGC [26] [27].

A large uninvestigated issue is the influence of social capital from a structural dimension perspective on the quality of reviews. Reviews are also a kind of UGC, so we have reasons to believe that the above conclusion is also applicable to online reviews, so this article examines how indegree and outdegree affect the review quality from a social capital perspective.

2.3. Review Quality

Reviews on a product have an important impact on sales [30], therefore studying quality of reviews has a great practical significance. The question “What kind of reviews is more helpful” is being considered from three aspects: firstly, content of review: analyzing content text of reviews through the length of reviews [12], readability, spelling error [14], emotional words,
et al. [38] explained that reviewers are the source of reviews. Whether a reviewer is trustworthy directly affects the quality of reviews. Past research has proposed reviewer information disclosure [31], rank [32], experience [30], the frequency that a reviewer access online community and so on. Pan et al. [6] have shown that the impact of review characteristics, product type and reviewer features on the helpfulness of online reviews and examined it with empirical analysis. Wu et al. [11] found that reviews with higher readability, less spelling errors, appropriate length and moderate emotion are more acceptable to consumers. Wu et al. [11] fully take the impact of information overload theory into account in the establishment of the model. In addition, publish time, consistency among reviews and description of products on the website also influence helpfulness of online reviews [28].

In terms of content of reviews, we prefer a reviewer’s characteristics such as reviewer’s experience and his/her activity, which may affect the quality of reviews.

3. Hypotheses and Research Model

3.1. Social Networks

Groups are an important feature of online community. Each group has its own theme. Users form their different but overlapping social networks by joining groups. Singh et al. [33] have explained that for OSS, moderate external cohesion has a positive effect on the success of the software. This means that if participants in one software development have moderate relationships with users who do not take part in this software development, this software is easier to succeed. Based on the “Bridging ties” theory in social networks, the relationships between online users in the same group is the “Bridging ties”. Aghakhani et al. [34] present that social ties have positive impacts on acceptance of explicit and implicit eWOM in Facebook. Joining more groups means having more bridging ties. It seems that the more the better, but Koroleva et al. [35] have reported that although users prefer information from their stronger ties in a network, high overlap in their networks decreases information value. Joining more groups will increase overlap in their networks and decrease information value, so we consider that there may be an “inverted U” shape relationship between the number of groups that a reviewer participates and review quality. In the same group, group members are interested in the same hobby.

The more the number of members, the more relative information about the hobby they can get, so the quality of reviews will be higher. Based on previous research, we propose the following hypotheses:

H1: There is an “inverted U” shape relationship between the number of groups that a reviewer participates and review quality. It means moderate number of groups has a positive influence on review quality.

H2: The average number of members in groups that a reviewer participates has a positive influence on review quality.

3.2. Centralization

For the structural dimension of social capital, the most common measurement is centralization, including degree centrality, betweenness centrality, closeness centrality, etc [36]. Degree centrality is defined as the number of ties that a focal actor has. Mallapragada et al. [26] have shown that in the open source software development process, when a creator’s centrality increases, the time to complete software will be reduced. Yang et al. [29] have indicated that in the open source software development process, with the increase of the number of a participant’s fans, the time to downloaded software will be increased. Pahlke [37] presented that in the knowledge exchange process, along with the increase of a participant’s centrality, the rate of helpful answers got higher. Susarla [28] examined that in a video sharing site, the more followers a video creator has, the easier the video will be spread and diffused. Peng et al. [38] explained that network centrality of current adopters fosters technology adoption. However, if a reviewer follows too many people, there will be people who agree with him and disagree with him, there will also be people who is right and not right, thus affecting the reviewer’s review quality. Based on previous research, we propose the following hypotheses:

H3: There is an “inverted U” shape relationship between the number of people that a reviewer follows and review quality. It means moderate number of people that a reviewer follows has a positive influence on the review quality.

H4: The number of a reviewer’s followers has a positive influence on the review quality.

3.3. Reviewer Characteristics

When review quality is not easy to estimate, review readers tend to consider reviewers’ credibility [39]. Trust development in different online collaboration groups does not follow a particular pattern [40].
Forman [41] examined that the richer relevant information a reviewer disclosed about himself, the more likely reviewers are to increase their own credibility. Bettina [42] has presented that reviewer’s experiences have great impact on his/her credibility. Ghose et al. [30] have tested that reviewer’s experiences have a positive influence on review quality, while some scholars have got different result. So we assume that there is an “inverted U” shape relationship between the number of reviews that a reviewer has post and review quality because if the number of reviews is too high for one reviewer, it is easy to make people feel that the quality of each review is not high since the energy of each person is limited. Sun et al. [43] have argued that review frequency has a positive effect on a reviewer’s credibility. In this paper we do not consider information disclosure and reviewers’ rank because of data collection. There is no uniform method to measure the degree of information disclosure and there is only Amazon has reviewers’ rank. Based on previous research, we propose the following hypotheses:

H5: There is an “inverted U” shape relationship between the number of reviews that a reviewer has post and review quality. It means moderate number of reviews that a reviewer has post has a positive influence on review quality.

H6: The interval days between current and last access time of a reviewer has a negative effect on review quality.

3.4. Review Helpfulness

Review helpfulness is the primary indicator of quality of reviews. It is the ratio of the helpful votes and total votes, higher proportion of helpful votes hints a more great influence on consumer purchasing decisions [12]. So, in this paper, we use review helpfulness ratio to measure review quality.

3.5. Control Variable

**Review Time:** Helpfulness of online reviews significantly depends on review time. Reviews posted earlier can get more votes than later posts. Many studies have shown that posting time is a key dimension of review quality [44].

**Total Votes:** Since helpfulness of online reviews is a relative number, some important information will be lost in the computing process. Such as “5 out of 8 people found the review helpful” may have a different explanation than “50 out of 80 people found the review helpful”. Many scholars have regard total votes as control variable [12].

**Similarity of Text Content:** Similarity of text content is the similarity between content of a review of a product and content of all the review of the same product. To compute its value, first of all, reviews are divided into many words, then get these words’ frequency in each review. Secondly, find out the weight of each word of all reviews through TF-IDF(Term Frequency–Inverse Document Frequency) Scheme and form a weight vector. Thirdly, use cosine similarity to calculate the similarity between the weight vector of each review and the weight vector of all reviews. Cao et al. [3] use many data mining techniques to analyze online review content. In a large number of reviews, a trend can be found that due to limitations of product itself, content of reviews will come to some keywords which respect characteristics of product. Those reviews which have more such keywords will reveal product better and can have more possibility to be high quality reviews. It means similarity of text content has a positive influence on review quality. This paper is to find out the relationship between review quality and reviewers’ characteristics (including reviewers’ social networks), so we regard the features of review text as a control variable.

3.6. Research Model

Our study focuses on the influence of reviewers’ social capital in an online community on review quality in China. Based on the discussion above, the theoretical model is showed in Figure 1.

![Figure 1. Research model](image)

4. Research Methodology

4.1. Data Collection
In this phase, we first used a crawler software to get 20 books’ reviews from one of Chinese largest online reading community (www.Douban.com). The number of reviews on each book is more than 30,000 (including those who just gave a star rating but did not write any text content), publish time of each book is ranges from January 2009 to December 2013, thus posting time of reviews can be controlled, here is what we get: reviewer’s name, content of review, posting time of review, total votes (TotalVotes), helpfulness (Helpfulness), the number of groups that a reviewer participates (GroupsNum), the number of members in groups that a reviewer participates, the number of people that a reviewer follows (FollowsNum), the number of people who follows a reviewer (FansNum), the number of reviews that a reviewer has post (Experience), the last access time of a reviewer. We then calculate the average number of members in groups (GroupsAve) that a reviewer participates and interval days between current and last access time of a reviewer (IntervalDay), and select reviews whose total votes is more than 1, and calculate the similarity of text content (ReviewSimilarity). The mean of similarity of text content is 0.366, the standard deviation is 0.125. In our final analysis, we chose reviews whose similarity of text content is in the range of (0.116, 0.617), which is two standard deviations from the mean, and the sample size is 1764. The definition of variable is in Table 1. The snapshot of a review on “Douban” is shown in Figure 2. The snapshot of a reviewer is shown in Figure 3.

<table>
<thead>
<tr>
<th>Table 1. Variable definitions</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>TotalVotes</td>
<td>Total votes of a review.</td>
</tr>
<tr>
<td>ReviewSimilarity</td>
<td>The similarity between content of a review and content of all the reviews.</td>
</tr>
<tr>
<td>Helpfulness</td>
<td>The ratio of helpful votes and total votes.</td>
</tr>
<tr>
<td>FollowsNum</td>
<td>The number of people that a reviewer follows.</td>
</tr>
<tr>
<td>FansNum</td>
<td>The number of people that follows a reviewer.</td>
</tr>
<tr>
<td>GroupsNum</td>
<td>The number of groups that a reviewer participates.</td>
</tr>
<tr>
<td>GroupsAve</td>
<td>The average number of members in groups that reviewer participates.</td>
</tr>
<tr>
<td>Experience</td>
<td>The number of reviews that a reviewer has post.</td>
</tr>
<tr>
<td>IntervalDay</td>
<td>The interval days between current and last access time of a reviewer.</td>
</tr>
</tbody>
</table>

4.2. Data Analysis

4.2.1. Descriptive statistics. The descriptive statistics for the variables in the full data set are included in Table 2.
The mean of total votes is about 38 and 83.2% people who voted think review is helpful. The mean of the number of groups that a reviewer participates is 84 and the mean of the average number of members is 51133, the minimum is 876. The mean of the number of reviews that a reviewer has post is 28, the mean of interval days between current and last access time of reviewer is 224.

4.2.2. The regression results. We use Tobit regression to analyze the model because of the feature of our dependent variable (helpfulness). Helpfulness is a ratio, so its value should be bounded in the range between 0 and 1. There is no difference between reviews that are just helpful and those that are excellent. The sample, however, has an inherent bias since not every review reader casts a vote. According to the hypotheses proposed, the regression equation can be constructed as follows:

$$\text{Helpfulness}^\% = \beta_0 + \beta_1 \text{Total Votes} + \beta_2 \ln(\text{Follows Num}) + \beta_3 \ln(\text{Follows Num})^2 + \beta_4 \text{Fans Num} + \beta_5 \text{Groups Num} + \beta_6 \text{Groups Ave} + \beta_7 \ln(\text{Experience}) + \beta_8 \ln(\text{Experience})^2 + \beta_9 \text{Interval Day}$$

Some variables in this regression model use log data to weaken heteroscedasticity and skewness that may appear in this model. Log data can avoid the influence caused by different units of measurement and extreme value [30]. In this step, we used Stata to analyze data, and measured goodness of fit with the Log-likelihood and pseudo R-square value. The regression result is shown in Table 3.

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Support Level</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1 Group Number</td>
<td>Not Supported</td>
<td>The average number of members in groups that a reviewer participates has a positive influence on review quality.</td>
</tr>
<tr>
<td>H2 Ave Groups’ Members</td>
<td>Supported</td>
<td>There is an “inverted U” shape relationship between the number of people that a reviewer follows and review quality.</td>
</tr>
<tr>
<td>H3 Follow Number</td>
<td>Supported</td>
<td>The number of people that follows a reviewer has a positive influence on the review quality.</td>
</tr>
<tr>
<td>H4 Follower Number</td>
<td>Supported</td>
<td>The interval days between current and last access time of a reviewer has a negative effect on review quality.</td>
</tr>
</tbody>
</table>

Log-likelihood (-13.018) and pseudo R-square value (0.604) indicate a good fit. The coefficient of the number of groups that a reviewer participates is positive and the coefficient of the square of the number of groups is negative, but the p-value is not significant, so H1 is not supported. The coefficient of the average number of members in groups is positive and the p-value (0.024) is significant, so H2 is supported. The coefficient of the number of people that a reviewer follows is positive and the coefficient of its square is negative, that means H3 is supported. The coefficient of the number of reviews that a reviewer has post is positive and the coefficient of its square is negative, but both of them are not significant, so H5 is not supported. The coefficient of interval days between current and the last access time of a reviewer is significantly negative, thus H6 is supported. The summary of results is presented in Table 4.

5. Discussion

The fact that H1 is not supported means that review readers are not affected by the number of groups a reviewer participates. H2 is supported, and it indicates
that review readers think the number of members in groups that a reviewer participates will influence the review quality. In our opinion, people in the same group talk about a specific issue. Information that members can get is accurate and precise. If the number of members in a group is large, the information and knowledge exchanged in the group will be more comprehensive on a specific issue. But if the number of groups that a reviewer participates is high, it will make it easy for review readers to feel that a reviewer’s knowledge is extensive but not professional, so the number of groups that a reviewer joins does not affect review quality.

H3 is supported, which illustrates that with people that a reviewer follows increasing, the information and knowledge that the reviewer obtains will increase, and their cognition will be more diversified, thus the review quality will increase too. However, when the number of people that a reviewer follows exceeds a threshold, the accuracy of information he receives will decline because he may follow someone who disagrees with himself/herself, thereby affecting the review quality. This indicates an “inverted-U” shape relationship between review quality and the number of people that a reviewer follows. H4 gets support too. If a reviewer has a lot of followers, review readers will consider that the reviewer is reliable, coupled with crowd mentality. Then it is easy to get a high quality review for the reviewer. In addition, reviewers who have a lot of followers will have a sense of responsibility and be serious to post review, therefore the review quality will be improved.

H5 is not supported, which presents the fact that the number of reviews a reviewer has post doesn’t have an influence on review quality. Some scholars have argued that the average of helpfulness votes of a reviewer’s previous reviews has a positive impact on quality of new reviews. It means that compared with the number of reviews posted by a reviewer, review readers are more concerned about the quality of their reviews. H6 is confirmed, and it indicates that the more active a reviewer is, the higher the review quality is. This is because when a reviewer visits an online review community frequently, he can learn more knowledge and it helps produce high quality reviews. Moreover, people with high exposure tend to be trusted by others. Another reason is that active reviewers will pay more attention to reviews. They want to be a good reviewer and they have time, thus the review quality will be improved.

6. Conclusion

6.1. General Conclusion

This study relies on social network theory and social capital theory, considers factors that affect the quality of reviews from reviewers’ network structure, social relationships and a reviewer’s own characteristics. We grab data from the Chinese largest book website to ensure the date is representative. Our study uses Tobit regression method to establish a theoretical model, explaining how the average number of members in a group that a reviewer participates, indegree and outdegree in social network, activity influence the quality of reviews. This research provides a new perspective to explore the effect of the complex mechanism on review quality and make contributions on collaboration research.

6.2. Research Implications

This paper is an important branch of collaboration research. Past researches have studied review quality from three aspects: content of reviews, star ratings, and features of reviewers. There is few research studied from reviewers’ social capital perspective. This paper has three theoretical contributions to fill several gaps in the literature relating online review quality. First of all, from social network perspective, it explores that the average number of members in the groups which a reviewer joins has a positive impact on review quality. Second, from social capital perspective, it illustrates that reviewers’ indegree has a positive effect on review quality and outdegree has an “inverted-U” shape relationship with review quality. Third, it develops that reviewers’ activity has a positive influence on quality of reviews. This study provides new perspectives for future research relating collaboration.

6.3. Practical Implications

For online community platforms, they can apply results of this model to improve online review effectively. First of all, online communities can recommend groups that have a large number of members to reviewers, let reviewers obtain more information and knowledge. Second, online communities can warn reviewers when they follow too many people since there is an “inverted U” shape relationship between the number of people a reviewer follows and review quality. Third, when online communities show content of reviews, they can also present reviewers’ information at the same time, such as the number of people that follows the reviewer, the average number of members in groups that the reviewer participates, the frequency that the reviewer visits online community, etc. At last, online
communities can sort reviews not only by helpfulness, but also by considering factors mentioned above to provide users with more optimized information. For product companies, finding out the helpful reviews can accurately find the needs of users, thus facilitate product upgrades. Companies can also seek the reviewer whose review quality is high to try out new products, it will achieve better propaganda effect. So this paper can promote the sound development of online reviews community in China and provide several good methods to let collaboration work well if regarding online communities platforms as leaders in collaborative projects.

6.4. Limitations & Further Research

This paper has two limitations. First, from structural dimension of social capital we only considered outdegree and indegree, without considering closeness centrality and betweenness centrality. As the primary method to measure structural dimension of social capital, closeness and betweenness can be carried in deeper analysis in future studies. Second, the object of the reviews in this paper is book which is experience product, we didn’t consider any search product. Future research can compare experience product and search product to find out the different effect on review quality caused by the same factors.

For future research, First of all, future research can regard closeness and betweenness as factors of reviewers’ social capital and take a group as a unit to build a social network, considering the influence of structural holes. Second, product can be classified into experience product and search product, besides, experience product can be divided into books, movies, and music etc. Third, adding time dimension, considering the dynamic evolution process between the reviewers’ social capital and quality of reviews. Fourth, considering other forms of collaboration from social capital perspective. Anyway, we hope this paper can let everyone have a deeper understanding of the influence of reviewers’ network characteristics and the features of reviewers on review quality.

7. References