MEASURING THE BUSINESS VALUE OF ONLINE SOCIAL MEDIA CONTENT FOR MARKETERS

Completed Research Paper

Zhijie Lin
Department of Information Systems
National University of Singapore
13 Computing Drive, Singapore 117417
linzhijie@comp.nus.edu.sg

Khim Yong Goh
Department of Information Systems
National University of Singapore
13 Computing Drive, Singapore 117417
gohky@comp.nus.edu.sg

Abstract

Marketers have been rushing to increase their social media marketing expenditures. However, the state of empirical research, business intelligence and analytics into the business value of social media marketing engagements have still lagged behind. Prior research about the impact of social media user-generated content on aggregate sales has overlooked the qualitative aspects. Moreover, due to the co-existence of consumers and marketers in social media, how these two roles generate business value has been understudied. Therefore, this study proposes the concept of social media marketer-generated content, and investigates the business value of user- and marketer-generated content, focusing on content sentiment and information. Ordinary least squares specification is employed to model sales performance. We find that the qualitative nature of social media content indeed has some business value. Specifically, we find a significant relationship between the richness of information embedded in both user- and marketer-generated content and firm sales performance.

Keywords: Social media, business value, user-generated content, marketer-generated content
Introduction

Social media sites have attracted enormous public’s interest. According to eMarketer.com, there were 70.6 million U.S. adult online social network users in 2009, which will reach 84.3 million in 2011 (Grau 2009). Businesses have been utilizing online social media for marketing, advertising and public relations. According to another report from eMarketer.com, the advertising revenue received by the popular social media site, Facebook.com, was $1.86 billion worldwide in 2010, and it predicted that this will reach $4.05 billion in 2011 (Williamson 2011). Despite the popular use of social media as a marketing channel, firms are still not clear about the business value of social media. As such, the state of empirical research, business intelligence and analytics into the business value of social media marketing engagements have still lagged behind.

Previous studies in user-generated content (UGC) (e.g., Chen and Xie 2008; Duan et al. 2008; Godes and Mayzlin 2004) tried to study the business value embedded in the UGC. However, these studies focused mainly on quantitative aspects (e.g., volume, rating) of UGC while overlooking its qualitative aspects. Consider the example of restaurant reviews, quantitative measures such as review volume and rating can only indicate how popular the restaurant is and consumers’ general evaluations of the restaurant and food. Consumers are not able to extract detailed product related information (e.g., taste of specific dishes, restaurant ambience, etc.) from these measures. Thus, consumers may still face many product uncertainties when they make purchase decisions. However, the textual content involves detailed product related information, and there are already some evidence suggesting that consumers actually read the textual content to seek more information about the product (e.g., Chevalier and Mayzlin 2006). Moreover, textual content is the main and even the only format in current mainstream social media (e.g., online news media, blog, company community), where researchers cannot directly capture ratings at all. Therefore, it is critical to study the business value of textual information in UGC, which has been understudied in past research.

Because of the development of technology, marketers are now able to interact with consumers in the social media context. As a typical example, marketers set up company fan pages on Facebook.com. They generate a large amount of contents to interact with consumers, aiming to drive them to purchase. In this study, in contrast to UGC generated by users (also referred to as “consumers” in this study), content generated by marketers is defined as marketer-generated content (MGC). Thus, consumers are making decisions under the influence of contents from both consumer peers (in the form of UGC) and marketers (in the form of MGC). As such, it might omit important information and conclude with imprecise results if MGC impact is overlooked. Therefore, it is necessary to examine the roles of both UGC and MGC.

By addressing above gaps, this paper unveils the underlying business value of social media content from a new perspective, which is the qualitative nature of the contents, and further presents a comprehensive investigation of the impact of two types of social media contents (UGC and MGC) on aggregate sales. As such, this study provides valuable insights on how marketers can leverage business intelligence techniques and social media for business success. To address these gaps, this study attempts to answer the following research question:

*How and to what extent is a firm’s sales performance influenced by user-generated content and marketer-generated content in the social media context?*

Using data of social media interaction content from a Facebook fan page, and offline sales data of an apparel retailer, this study explores the impact of social media user-generated content and marketer-generated content on firm sales performance. Two focused factors, i.e., content sentiment and content information, are explored in this study.

We use ordinary least squares regression model for estimation. Our findings show that qualitative information embedded in social media content is correlated to changes in aggregate sales outcome. Specifically, the richness of information in both user- and marketer-generated content has a significant relationship with sales performance. However, we find no support for the relationship between content sentiment and sales.

In summary, this study has the following contributions. First, it sheds light on the business value of social media for marketers by investigating the impact of social media content on aggregate sales performance.
Second, it addresses the research gap that qualitative information has been overlooked in prior research on user-generated content and shows some evidence that the qualitative nature of social media content indeed has some business impact and is even more important than the quantitative nature. Third, it points out that prior studies did not observe the co-existence of consumers and marketers in the social media context, and this study comprehensively examines the impact of these two types of contents on aggregate sales outcome. Lastly, this study demonstrates a novel methodological approach for business intelligence that integrates text mining techniques and econometric methods for empirical research.

Literature Review and Hypotheses

Prior Literature

Previous studies about the business value of UGC have invariably focused on the quantitative measures (e.g., Chen et al. 2004; Chevalier and Mayzlin 2006; Clemons et al. 2006; Dellarocas et al. 2004; Liu 2006; Moon et al. 2010; Ye et al. 2009). Specifically, Liu (2006) investigated the impact of review volume and user ratings on box office revenue and concluded with the result that volume instead of rating is the dominating factor in predicting box office revenue. Ye et al. (2009) found supportive evidence using sales data from hotel room booking website and concluded that higher ratings significantly increase hotel room sales. Moon et al. (2010) studied the impact of movie ratings on box office revenue also using movie data. Similarly, they found that higher movie ratings can significantly increase movie revenue. Findings from these past studies show that the significant relationship between UGC quantitative measures and aggregate sales outcome has received strong support, although mixed and even conflicting results were reported.

Although many studies have documented the significant relationship between UGC quantitative measures and aggregate sales, they only incorporated the quantitative aspects but overlooked the qualitative nature. The prior study by Chevalier and Mayzlin (2006) however bucked this trend and shed some light on the significant relationship between the length of consumer reviews and aggregate sales, suggesting that consumers did read and respond to the written reviews. However, there still exists the paucity of research studying the qualitative aspects of social media content, which constitutes a research gap to be addressed in this study. On the other hand, due to marketers’ increasing engagement in social media, content generated by marketers co-exists with UGC. Thus consumers are concurrently exposed to both types of contents. However, the business value of contents generated by marketers, and how this business value is different from that of UGC, remain unsolved in existing studies.

Constructs and Hypotheses Development

Although many textual features are relevant, two factors are identified to be important from the literature, i.e., content sentiment and content information. Sentiment expressed in the content, positive, negative or indifferent, represents consumers’ general attitude towards the product/brand/firm. This may reflect their evaluations. Thus, sentiment can serve as a quality index for consumers to refer to, and eventually it may be correlated to changes in aggregate sales outcome. In this study, sentiment is defined as favorableness of the content. On the other hand, information expressed in the content plays another important role. Consumers are seeking more product information from the textual content (Chevalier and Mayzlin 2006) to minimize their uncertainties. Therefore, the amount of information may serve as another critical qualitative factor in reducing consumer uncertainties and consequently lead to changes in aggregate sales outcome. In this study, amount of information is defined as information richness of the content.

In sum, favorableness and information richness are employed in this study to explore the relationship between social media content (UGC and MGC) and aggregate sales performance.

Content Favorableness and Sales

Since consumers generate UGC to interact with consumers and marketers, we believe that sentiment embedded in consumers’ UGC is highly relevant to the product of interest. As such, sentiment represents consumers’ general attitude towards the product. If consumers are more satisfied with the product, they
Knowledge Management and Business Intelligence

will have more favorable sentiment, whereas they will have negative sentiment if they dislike the product. Therefore, favorableness of the sentiment can be interpreted as consumers’ general evaluations of the product.

There exist some studies about consumers’ product evaluations. Ratings indicate consumers’ general evaluations of the product. Higher ratings imply that consumers are more satisfied with the product, whereas low ratings may imply that they dislike it. The impact of consumer evaluations on aggregate sales outcome has received strong support from past studies. For instance, Dellarocas et al. (2004) focused on the motion picture industry and developed a model based on the Bass model (Bass 1969). They reported that their predictive model forecasts revenue accurately after incorporating numerical consumer ratings. Zhu and Zhang (2010) examined the influence of online consumer reviews on product sales using data from the video game industry. The findings indicate that higher consumer ratings result in higher product sales, and the rating impact is more influential for less popular games and games whose players have greater Internet experience.

Based on above discussion, since favorableness of the UGC reflects consumers’ product evaluations, we posit that there exist a significant relationship between UGC favorableness and aggregate sales outcome:

Hypothesis 1a (H1a): Favorableness of social media UGC has a positive relationship with aggregate sales.

Marketers generate marketing and advertising messages to interact with the consumers, aiming to increase sales. As MGC is generated by marketers, we believe that MGC includes marketing and advertising information, which can drive consumer purchase (Manchanda et al. 2006; Zhang and Wedel 2009), and the embedded persuasion was found to be one of the major underlying factors driving consumer purchase (Meyers-Levy and Malaviya 1999; Russo and Chaxel 2010; Wu et al. 2009). As such, MGC sentiment can be interpreted as marketers’ persuasion. Thus, from the consumers’ perspective, MGC favorableness implies marketers’ strength of persuasion. When consumers perceive higher level of MGC favorableness, they may be more likely to be persuaded to purchase. Therefore, we posit that MGC favorableness may have a positive relationship with aggregate sales outcome:

Hypothesis 1b (H1b): Favorableness of social media MGC has a positive relationship with aggregate sales.

Content Information Richness and Sales

Product quality can be observed perhaps only after purchase, especially for experience goods such as apparels, foods, etc. This causes a limitation of quality information to consumers, which has been documented in previous studies (e.g., Hey and McKenna 1981; Narayan et al. 2007; Nelson 1970). Thus, consumers often need to make purchase decisions under risks. However, consumers are averse to losses (Kahneman and Tversky 1979), thus they are seeking more product information to reduce their quality uncertainties. If they have more information about product quality, they may be more confident about their purchase decisions. As a result, we would observe that they are more likely to purchase.

Consumers generate UGC to interact with peers and marketers. They express their personal perceptions or experiences about the product. Thus, UGC involves product related information. Although UGC may not necessary indicate product quality in a direct manner, it still can reflect the quality based on the product related information. Thus, consumers can infer the quality of the product based on this information. As such, the more information involved in UGC, the better it can reflect the quality of the product, and the fewer quality uncertainties consumer will perceive. Consequently, consumers may be more likely to purchase. Previous studies have shown some support for this relationship. For instance, Forman et al. (2008) examined the relationship between online reviews and aggregate sales by including the information about reviewer identity, and they reported that this disclosure of extra information significantly increases sales performance. Therefore, based on above discussion, this study postulates that higher UGC information richness will drive more consumers to purchase, which leads to better sales performance:

Hypothesis 2a (H2a): Information richness of social media UGC has a positive relationship with aggregate sales.
Similar to the case of UGC, MGC also involves product related information since marketers often need to introduce new products and launch marketing campaigns. As such, consumers can also extract quality information from the product information embedded in MGC. The more information provided by the MGC, the better MGC can reduce consumers’ quality uncertainties. Consequently, we would observe consumers are more likely to purchase, which leads to better aggregate sales outcome. Therefore, we posit that information richness of MGC may also have a positive relationship with aggregate sales outcome:

Hypothesis 2b (H2b): Information richness of social media MGC has a positive relationship with aggregate sales.

Research Methodology

Research Model

![Research Model Diagram]

Figure 1 presents our proposed research model with an illustration of the relationships between the model constructs. We focus on two textual features (favorableness and information richness) of social media UGC and MGC. We study the business value of these two types of contents by investigating the relationships between the textual features of social media content and firms’ aggregate sales performance. The control variables are gathered from the outcome of our literature review and from the information which is available to us in our data. We propose four hypotheses relating the constructs studied in our research model. Using this research model, we seek to gain a holistic understanding of the business value of social media content.

Research Context

We conduct our empirical study in the context of user interactions on the FFS\(^1\) apparel retailer’s fan page community on Facebook.com. Figure 2 presents an edited screen shot of this fan page community. FFS retailer set up this community to serve as a platform for user interactions. Online users can simply “like” this fan page community to engage as community members (also referred to as “consumers” in this study). Thus, there are two types of users in this fan page community, i.e., consumers and the marketer.

---

\(^1\) Due to confidentiality agreements, we are not able to reveal the identity of the retailer.
Knowledge Management and Business Intelligence

(FFS Retailer). The marketer and consumers generate a large amount of MGC and UGC, respectively. Content (posts and comments) generated by the marketer is referred to as MGC, whereas content (posts and comments) generated by consumers is referred to as UGC. The marketer generates MGC to launch marketing campaigns and interact with the consumers, whereas consumers generate UGC to interact with the marketer and other consumer peers to share their attitudes, evaluations, etc.

FFS fan page community presents an ideal environment for this study. First, contents in this community serve as a representative example of UGC and MGC. Second, there exist extensive user interactions, which provide substantial information for our study. Third, almost all the UGC and MGC in this community are in textual format (only a small number of pictures and videos). Thus this satisfies our research focus on the qualitative aspects of social media content. Lastly, according to FFS retailer, Facebook is the only social media channel they use to connect consumers, thus this provides a clean and unambiguous setting to investigate the impact of social media content on sales.

![Image: FFS Fan Page Community](image)

**Empirical Model Specification**

We observe social media interactions and offline sales at daily level. The dependent variable is daily aggregate sales revenue (\(SALES\)). The independent variables we focus on are the UGC and MGC factors, which include UGC favorableness (\(UF\)), UGC information richness (\(UI\)), MGC favorableness (\(MF\)) and MGC information richness (\(MI\)). In addition, we include control variables such as firm’s promotional activity, a binary indicator of the existence of promotional event at each day (\(PROM\)), and traditional quantitative factors UGC volume (\(UV\)) (daily amount), MGC volume (\(MV\)) (daily amount). We employ ordinary least squares (OLS) regression model for estimation.

Our choice of the daily analysis was due to the following practical reasons: (1) daily level analysis can provide more sample observations compared with weekly/monthly level analysis; (2) some consumers actually made more than one purchase incidence in the same week, thus daily level analysis can give us more detailed information at consumer transaction level; and (3) we tried to avoid aggregation bias when aggregating the daily level data to weekly or monthly level (Christen et al. 1997; Theil 1971).
An important issue in this study is that consumers purchase in the offline context. In other words, it is more logical and reasonable to assume that current contents will impact on future sales, instead of current sales. This is different from online shopping, where consumers can purchase immediately after they refer to some online information in the same time period (e.g., day). Therefore, we introduce the lagged terms of all social media content factors, and investigate the relationship between past contents (at day \( t-j \)) and current sales (at day \( t \)). Overall, this yields the following estimation equation:

\[
SALES_t = \alpha + \sum_{j=1}^{J} \beta_j UF_{t-j} + \sum_{j=1}^{J} \beta_j MF_{t-j} + \sum_{j=1}^{J} \beta_j UI_{t-j} + \sum_{j=1}^{J} \beta_j M_{t-j} + \sum_{j=1}^{J} \beta_j UV_{t-j} + \sum_{j=1}^{J} \beta_j MV_{t-j} + \beta PROM_t + \epsilon,
\]

where \( \alpha \) denotes the unobserved firm specific characteristics. \( \beta \)s is a set of coefficients. \( \epsilon \) indicates the random error term. We consider the impact of contents in the past \( J \) days on current sales.

**Data Description**

**Data Collection**

We collected data from two sources. First, we designed Java codes based on Facebook API (application programming interface) to retrieve user interaction data from FFS retailer fan page community on Facebook.com. These data include contents (posts and comments) generated by both consumers and the marketer. Second, FFS retailer provided us with the aggregate sales data of existing consumers in their loyalty card program. In addition, the company released to us comprehensive information about each and every marketing promotion in their marketing calendar. Therefore, we are able to construct all variables in our study and test our hypotheses by linking up these two parts of data.

We collected the data from September 8, 2009 to June 30, 2011. However, we found some missing days in the daily transaction data. Therefore, our data eventually included 643 days for observation.

**Variables and Measures**

We employ a novel approach to analyze the textual data for econometric analysis. Specifically, we employ a business intelligence tool (a commercial text mining tool), SPSS Clementine, to analyze the textual contents of UGC and MGC. This tool includes a large library of words, phrases from various domains or disciplines. Given a piece of textual content (can be only a word to a lengthy article), this tool will decompose the content into words and phrases based on its large library, and then extract concepts from these words and phrases. For each concept extracted, this tool will assign a type to the concept. Types are used to indicate the sentiment nature (positive, negative, indifferent) of the concepts.

Figure 3 illustrates the text analysis results. We input three pieces of text examples. The first text from the marketer (ID=1) is "Come to the outlet near City Hall! From today till 1st August, enjoy your 30% discount!". The other two texts from consumers (ID=2, ID=3) are: "That's great! I always love your jeans, make me look so good!", "I hate that place, that's really far away from my home and so dirty there". As indicated in Figure 3, "ID" shows the index of each piece of text. "Matched Text" shows the original text. "Concept1" and "Concept2" are the extracted concepts (indicated by brackets <*> in the original text). "Type1" and "Type2" are the corresponding sentiment nature of each extracted concept. Each row will display two extracted concepts at most. "Null" indicates that no concept is extracted. Positive (negative) sentiment can be identified by type value with "Positive" ("Negative") or "Positive Qualifier" ("Negative Qualifier"), and all other types are considered as indifferent in this study. For instance, five concepts are identified from the second text (ID=2), with three of them positive, none of them negative, and two of them indifferent.

Therefore, as the number of concepts can reflect the richness of information, and the types of the extracted concepts can reflect the sentiment nature, our measures of UGC and MGC factors are directly derived from these text mining results. Specifically, content information richness is measured by the number of concepts extracted, and content favorableness is measured by net positivity (the number of positive concepts minus the number of negative concepts).
Additionally, the dependent variable, **SALES**, is measured by total sales revenue (in thousands of dollars) each day. UGC/MGC volume (**UV/MV**) is measured by the number of entries of UGC/MGC generated each day. Promotional information is captured from the information about each and every marketing promotion in their marketing calendar. We use a binary indicator, **PROM**, to capture the existence of promotional event in each day. Table 1 provides a summary of the measures for all variables used in our study. Tables 2 and 3 present the descriptive statistics and correlations of all variables.

![Figure 3. Text Analysis Results](image)

### Table 1. Summary of Measures

<table>
<thead>
<tr>
<th>Construct</th>
<th>Variable</th>
<th>Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sales revenue</td>
<td><strong>SALES</strong></td>
<td>Total sales revenue each day (in thousands of dollars)</td>
</tr>
<tr>
<td>UGC favorableness</td>
<td><strong>UF</strong></td>
<td>The number of positive concepts minus the number of negative concepts, for all UGC generated each day</td>
</tr>
<tr>
<td>MGC favorableness</td>
<td><strong>MF</strong></td>
<td>The number of positive concepts minus the number of negative concepts, for all MGC generated each day</td>
</tr>
<tr>
<td>UGC information richness</td>
<td><strong>UI</strong></td>
<td>The number of concepts in all UGC generated each day</td>
</tr>
<tr>
<td>MGC information richness</td>
<td><strong>MI</strong></td>
<td>The number of concepts in all MGC generated each day</td>
</tr>
<tr>
<td>UGC volume</td>
<td><strong>UV</strong></td>
<td>The number of entries of UGC generated each day</td>
</tr>
<tr>
<td>MGC volume</td>
<td><strong>MV</strong></td>
<td>The number of entries of MGC generated each day</td>
</tr>
<tr>
<td>Promotion</td>
<td><strong>PROM</strong></td>
<td>Existence of promotional event each day (0/1 indicator)</td>
</tr>
</tbody>
</table>

### Table 2. Descriptive Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SALES</strong></td>
<td>643</td>
<td>10.96</td>
<td>8.75</td>
<td>0.28</td>
<td>49.18</td>
</tr>
<tr>
<td><strong>UF</strong></td>
<td>643</td>
<td>0.15</td>
<td>0.94</td>
<td>-4.00</td>
<td>7.00</td>
</tr>
<tr>
<td><strong>MF</strong></td>
<td>643</td>
<td>0.35</td>
<td>1.30</td>
<td>-3.00</td>
<td>12.00</td>
</tr>
<tr>
<td><strong>UI</strong></td>
<td>643</td>
<td>2.61</td>
<td>6.73</td>
<td>0.00</td>
<td>62.00</td>
</tr>
<tr>
<td><strong>MI</strong></td>
<td>643</td>
<td>3.92</td>
<td>8.37</td>
<td>0.00</td>
<td>59.00</td>
</tr>
<tr>
<td><strong>UV</strong></td>
<td>643</td>
<td>0.90</td>
<td>2.72</td>
<td>0.00</td>
<td>44.00</td>
</tr>
<tr>
<td><strong>MV</strong></td>
<td>643</td>
<td>0.57</td>
<td>1.11</td>
<td>0.00</td>
<td>8.00</td>
</tr>
<tr>
<td><strong>PROM</strong></td>
<td>643</td>
<td>0.68</td>
<td>0.47</td>
<td>0.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>
### Table 3. Correlations

<table>
<thead>
<tr>
<th></th>
<th>SALES</th>
<th>UF</th>
<th>MF</th>
<th>UI</th>
<th>MI</th>
<th>UV</th>
<th>MV</th>
<th>PROM</th>
</tr>
</thead>
<tbody>
<tr>
<td>SALES</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UF</td>
<td>0.07</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MF</td>
<td>0.11</td>
<td>0.20</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UI</td>
<td>0.16</td>
<td>0.44</td>
<td>0.35</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MI</td>
<td>0.11</td>
<td>0.29</td>
<td>0.64</td>
<td>0.48</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UV</td>
<td>0.13</td>
<td>0.28</td>
<td>0.38</td>
<td>0.78</td>
<td>0.51</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MV</td>
<td>0.07</td>
<td>0.30</td>
<td>0.56</td>
<td>0.50</td>
<td>0.89</td>
<td>0.51</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>PROM</td>
<td>0.38</td>
<td>0.08</td>
<td>0.06</td>
<td>0.04</td>
<td>0.08</td>
<td>0.02</td>
<td>0.04</td>
<td>1.00</td>
</tr>
</tbody>
</table>

### Data Analysis

#### Empirical Results

In order to alleviate the concern of potential multicollinearity issue, we mean-centered all UGC and MGC factors (UF, MF, UI, MI, UV, MV) before our formal estimation. Another concern is about the potential existence of heteroskedasticity. As such, we reported robust standard errors to ensure that our estimated standard errors are unbiased. The detailed estimation process is summarized in Table 4.

We started with the one-day lagged terms of UGC and MGC factors. The estimation results using these one-day lagged UGC and MGC terms are shown in Model 1. Specifically, the coefficient of UGC information richness (UI\(_{t-1}\)), 0.28 (±0.10) (p<0.01), was positive and statistically significant. Similarly, the coefficient of MGC information richness (MI\(_{t-1}\)), 0.20 (±0.09) (p<0.01), was positive and statistically significant. However, the coefficients of UGC favorableness (UF\(_{t-1}\)) and MGC favorableness (MF\(_{t-1}\)) were insignificant according to the results.

We then replaced the one-day lagged terms with two-day and three-day lagged terms, and conducted similar estimations using these lagged variables. The results of these two estimations are accordingly shown in Model 2 and Model 3. As indicated, these two models yielded qualitatively similar results that only UGC information richness and MGC information richness were found to be positively correlated to changes in sales revenue. Specifically in Model 2, the coefficient of UGC information richness (UI\(_{t-2}\)) was 0.23 (±0.10) (p<0.05), and the coefficient of MGC information richness (MI\(_{t-2}\)) was 0.28 (±0.09) (p<0.01). Similarly in Model 3, the coefficients of UI\(_{t-3}\) and MI\(_{t-3}\) were 0.32 (±0.10) (p<0.01) and 0.19 (±0.09) (p<0.05), respectively.

Although not reported, we further estimated using four-day lagged terms. However, the yielded results started to show insignificant relationship between UGC/MGC information richness and sales revenue (UGC/MGC favorableness remained insignificant at this step). As such, we considered the lagged effect of UGC and MGC factors on sales up to three days.

We lastly estimated the cumulative impact of UGC and MGC factors on sales. The results are shown in Model 4, which are qualitatively similar to the results from previous models. The only difference is that the coefficient of MI\(_{t-1}\) became insignificant (the coefficients of MI\(_{t-2}\) and MI\(_{t-3}\) remained positively significant). However, this difference did not change our conclusion that MGC information richness has a positive relationship with sales.

As to the control variables, we included traditional quantitative measures of the content (UV, MV) and the firm’s marketing activity (PROM). As the results indicated, PROM was consistently significant in all estimated models, whereas UV and MV were generally insignificant. This insignificant relationship suggested that the qualitative impact might be more important than the quantitative impact of social media content.
Table 4. Estimation Results (OLS Model)

<table>
<thead>
<tr>
<th>Construct</th>
<th>Variable</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>UGC favorableness</td>
<td>UF&lt;sub&gt;t-1&lt;/sub&gt;</td>
<td>-0.18 (0.41)</td>
<td>-0.46 (0.37)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>UF&lt;sub&gt;t-2&lt;/sub&gt;</td>
<td>-0.40 (0.44)</td>
<td>-0.54 (0.40)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>UF&lt;sub&gt;t-3&lt;/sub&gt;</td>
<td></td>
<td>-1.05 (0.56)</td>
<td>-0.94 (0.51)</td>
<td></td>
</tr>
<tr>
<td>MGC favorableness</td>
<td>MF&lt;sub&gt;t-1&lt;/sub&gt;</td>
<td>0.46 (0.41)</td>
<td></td>
<td>0.51 (0.33)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MF&lt;sub&gt;t-2&lt;/sub&gt;</td>
<td>0.56 (0.41)</td>
<td></td>
<td>0.45 (0.38)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MF&lt;sub&gt;t-3&lt;/sub&gt;</td>
<td></td>
<td>0.18 (0.36)</td>
<td>0.11 (0.37)</td>
<td></td>
</tr>
<tr>
<td>UGC information richness</td>
<td>UI&lt;sub&gt;t-1&lt;/sub&gt;</td>
<td>0.28** (0.10)</td>
<td></td>
<td>0.21** (0.08)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>UI&lt;sub&gt;t-2&lt;/sub&gt;</td>
<td>0.23* (0.10)</td>
<td></td>
<td>0.21** (0.08)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>UI&lt;sub&gt;t-3&lt;/sub&gt;</td>
<td>0.32** (0.10)</td>
<td>0.30** (0.08)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MGC information richness</td>
<td>MI&lt;sub&gt;t-1&lt;/sub&gt;</td>
<td>0.20** (0.09)</td>
<td></td>
<td>0.14 (0.09)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MI&lt;sub&gt;t-2&lt;/sub&gt;</td>
<td>0.28** (0.09)</td>
<td></td>
<td>0.24** (0.09)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MI&lt;sub&gt;t-3&lt;/sub&gt;</td>
<td>0.19* (0.09)</td>
<td>0.18* (0.08)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UGC volume</td>
<td>UV&lt;sub&gt;t-1&lt;/sub&gt;</td>
<td>-0.10 (0.25)</td>
<td>-0.17 (0.13)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>UV&lt;sub&gt;t-2&lt;/sub&gt;</td>
<td>0.27 (0.30)</td>
<td></td>
<td>0.11 (0.22)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>UV&lt;sub&gt;t-3&lt;/sub&gt;</td>
<td>-0.08 (0.22)</td>
<td>-0.28 (0.17)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MGC volume</td>
<td>MV&lt;sub&gt;t-1&lt;/sub&gt;</td>
<td>-0.94 (0.64)</td>
<td>-0.84 (0.59)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MV&lt;sub&gt;t-2&lt;/sub&gt;</td>
<td>-1.16 (0.67)</td>
<td>-1.14* (0.56)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MV&lt;sub&gt;t-3&lt;/sub&gt;</td>
<td>-0.22 (0.68)</td>
<td>-0.13 (0.57)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Promotion</td>
<td>PROM&lt;sub&gt;t&lt;/sub&gt;</td>
<td>6.61*** (0.56)</td>
<td>6.16*** (0.54)</td>
<td>6.70*** (0.56)</td>
<td>6.10*** (0.52)</td>
</tr>
<tr>
<td>Constant</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of observations</td>
<td>628</td>
<td>626</td>
<td>624</td>
<td>603</td>
<td></td>
</tr>
<tr>
<td>R&lt;sup&gt;2&lt;/sup&gt;</td>
<td>0.22</td>
<td>0.28</td>
<td>0.24</td>
<td>0.39</td>
<td></td>
</tr>
</tbody>
</table>

Notes: *p<0.05; **p<0.01; ***p<0.001; robust standard errors in parentheses.

Based on above estimation results, our two hypotheses H2a (UGC information richness and sales) and H2b (MGC information richness and sales) were supported. However, we could not find support for H1a (UGC favorableness and sales) and H1b (MGC favorableness and sales). The plausible reason for the insignificant relationship might be that, over the years, consumers have developed a general tendency of disbelief or skepticism towards persuasion messages. This can be related to the literature on information and persuasion effects of advertisements. For instance, Ackerberg (2001) empirically distinguished the informative and persuasive effects of advertisements, and concluded that advertisements only influence consumer behavior by informing them about product characteristics, but not by creating prestige.
Discussion and Implications

The main aim of our study is to discover the business value of social media by empirically investigating the relationship between social media user- and marketer-generated content and firm sales performance. We conduct our analysis in the context of an online community set up by an apparel retailer. Using business intelligence techniques, we focused on two qualitative aspects (content favorableness and content information richness) of social media content, and proposed hypotheses regarding the impact of these qualitative factors on firm sales performance.

From our results, the hypotheses regarding the relationship between UGC and MGC information richness and sales were supported. Specifically, in a social media context where consumers and marketers engage with each other, the more information embedded in the content generated by consumers, the better sales performance firms will have. Similarly, the more information embedded in the content generated by marketers, the better sales performance firms will benefit.

This study provides some theoretical implications. First, this study discovers the business value of social media by showing the significant relationship between the information in social media content and firms’ sales performance. Thus, this study contributes to the IS literature by revealing the business value of social media.

Second, this study addresses the research gap that qualitative information has been overlooked in prior research on user-generated content. It shows some evidence that the qualitative nature of online social media content has some business value. Thus, this is one of the very few studies that have conducted and reported a comprehensive investigation of both quantitative and qualitative aspects of social media content. In addition, our results show that, after incorporating the qualitative aspects into the model, the quantitative impact became insignificant. This suggests that the qualitative nature of social media content might be more important than the quantitative nature. In other words, our study provides some new findings and insights, which are different from past studies.

Third, this study addresses another research gap that prior studies did not observe the co-existence of influences from consumer peers and marketers in the social media context. Thus, this study contributes to the IS and marketing literature by proposing the concept of “marketer-generated content”, in contrast to user-generated content, and comprehensively examines the impact of these two different types of information. Our findings suggest that under the co-existence of consumer peers and marketers, consumers will be affected by information from both of them. Therefore, this study extends the literature on user-generated content by outlining another important factor (marketer-generated content), which also has an impact on some business outcomes.

Fourth, this study is one among very few that integrates business intelligence techniques (i.e., text mining) and econometric methods to evaluate the business value of social media content. Facilitated by business intelligence techniques, business related data, especially qualitative data, can be processed and analyzed. As such, qualitative data can be linked to quantitative data for business analysis. Therefore, this study provides and especially demonstrates a novel methodological approach for empirical research.

This study offers a few important practical implications as well. First, this study has revealed the business value of social media. Therefore, apart from doing traditional offline marketing activities (e.g., promotion, discount, etc.) to drive sales, companies may also get better off to set up some social media sites to engage consumers. Our findings suggest that firms can benefit sales growth from social media engagement.

Second, our results show that information embedded in social media content from both consumers and marketers will affect firm sales performance. As such, on the one hand, marketers can generate more informative content to increase the richness level of marketer-generated content. For instance, marketers can share greater details in brand history, product features, and even the preparation process of some marketing activities, etc. On the other hand, marketers can also encourage consumer to interact with each other. For instance, marketers can offer some attractive coupons to consumers who are active in online interactions.

Third, this study has demonstrated the use of business intelligence techniques to analyze business related data, and shown significant relationship between social media content and sales. Based on this, managers
can utilize similar business intelligence techniques to analyze various business data to extract valuable information for business success.

**Conclusion**

While this research has found several notable new findings, we acknowledge some limitations of this study. First, the sample size in this study is relatively small. We only have available sales data from one company, although we are able to extract social media content from a large number of online communities. Second, besides textual content, there are some other formats of contents in our research context, such as pictures. However, these pictures were posted by the marketer, including some e-coupons or other forms of promotional messages, which at the same time appeared in the textual content captured in our sample. Therefore, we are able to partially observe such information. Lastly, the introduction of lagged variables might cause multicollinearity issue. However, this introduction was necessary due to our research context, and we already performed centralization to alleviate this concern.

**References**


