From Buzz to Bucks

The Impact of Social Media Opinions on the Locus of Innovation

Zhang, Wenping; Kang, Lele; Jiang, Qiqi ; Pei, Lei

Document Version
Accepted author manuscript

Published in:
Electronic Commerce Research and Applications

DOI:
10.1016/j.elerap.2018.04.004

Publication date:
2018

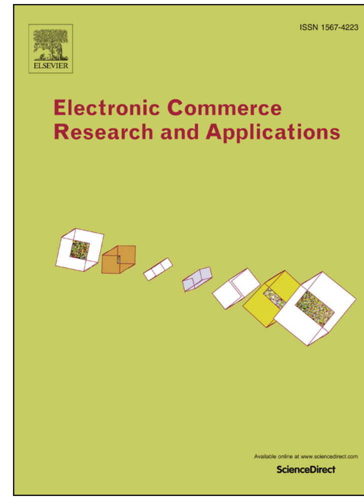
From Buzz to Bucks: The Impact of Social Media Opinions on the Locus of Innovation

Wenping Zhang, Lele Kang, Qiqi Jiang, and Lei Pei

Journal article (Accepted manuscript*)

Please cite this article as:
Zhang, W., Kang, L., Jiang, Q., & Pei, L. (2018). From Buzz to Bucks: The Impact of Social Media Opinions on the Locus of Innovation. *Electronic Commerce Research and Applications*, *30*, 125-137. https://doi.org/10.1016/j.elerap.2018.04.004

DOI: 10.1016/j.elerap.2018.04.004



From buzz to bucks: the impact of social media opinions on the locus of innovation

Wenping Zhang, Lele Kang, Qiqi Jiang, Lei Pei

PII: S1567-4223(18)30039-5
DOI: https://doi.org/10.1016/j.elerap.2018.04.004
Reference: ELERAP 787

To appear in: *Electronic Commerce Research and Applications*

Received Date: 7 October 2017
Revised Date: 26 April 2018
Accepted Date: 26 April 2018

Please cite this article as: W. Zhang, L. Kang, Q. Jiang, L. Pei, From buzz to bucks: the impact of social media opinions on the locus of innovation, *Electronic Commerce Research and Applications* (2018), doi: https://doi.org/10.1016/j.elerap.2018.04.004

FROM BUZZ TO BUCKS: THE IMPACT OF SOCIAL MEDIA OPINIONS ON THE LOCUS OF INNOVATION

Wenping Zhang
Renmin University
Beijing, China
wpzhang@ruc.edu.cn

Lele Kang (contact author)
Department of Information Management Science
Nanjing University
Nanjing, China
lelekang@nju.edu.cn

Qiqi Jiang
Copenhagen Business School
Copenhagen, Denmark
qj.digi@cbs.dk

Lei Pei
Department of Information Management Science
Nangjing University
plei@nju.edu.cn

Last revised: April 27, 2018

ABSTRACT

Social media influences business practices such as innovation activities. This study is anchored in the theoretical paradigm of strategic information technology (IT) alignment and the literature on social media and its business value. We examine how the valence and volume of *user-generated content* (UGC) from social media influence firm-level innovation strategies. Based on an analysis of five years of panel data for 886 listed firms and their 6.2 million relevant microblogs, we observed three interesting results. First, the valence of social media UGC has a U-shaped relationship with firms' innovation investments; and compared with neutral UGC, both negative and positive content lead to more innovation investment. Second, we found that this curvilinear relationship is mitigated as the volume of UGC increases. Third, we verified that increasing innovation investment improves firm performance. We then examined these findings using a series of strict robustness checks and discussed our study's contribution to theory and practice.

Keywords: Firm performance; innovation investment; microblogs; panel data; sentiment; social media; strategic IT alignment; user-generated content (UGC); valence; volume

1. INTRODUCTION

In the United States, social media use accounts for nearly three quarters of users' online time (Casey 2017); and this proportion has increased threefold since 2010 (Gallaugher and Ransbotham 2010). A similar situation exists in China, which is the largest Internet market based on the number of users (Miniwatts Marketing Group 2017). In a recent financial report, Weibo (2017), which is the dominant player in Chinese microblogging, indicated that it had 361 million monthly active users, which surpasses even Twitter, the target of Weibo's early imitation. Moreover, this large number of users helped Weibo to generate net revenue of USD 253.4 million by the end of June 2017.

Since its emergence, social media'has exerted tremendous influence on various business activities, including advertising (Chen et al. 2014), promotions (Kumar and Rajan 2012), and innovation practices (Ernst and Brem 2017; Zheng and Zheng 2014). Specifically, user-generated content (UGC) from social media is instrumental in helping firms make various strategic decisions (Kaplan and Haenlein 2010). For instance, firms that gain a better understanding of their customers by analyzing social media UGC outperform their rivals in marketing practices (He et al. 2013); firms can also strategically integrate UGC from leading users into their own social media as part of the innovation process (Ernst and Brem 2017). To this end, we found that social media, particularly UGC, is an important information technology (IT) resource that helps firms outperform their competitors; this echoes the notion of strategic IT alignment in the prior literature (Coltman et al. 2015). The activities of firms are especially influenced when the UGC on social media relates to the firms' products or services.

The theory of strategic IT alignment is an attempt to understand how firms strategically translate their deployment of IT to strengthen their business performance (Bergeron et al. 2004). Researchers have suggested that IT investment cannot contribute to improved firm performance if the alignment between business and IT is not harmoniously managed (Avison et al. 2004). In other words, a misfit in strategic alignment results in a waste of IT investment and negative consequences for business performance, such as a failures when implementing enterprise resource planning (ERP) (Amid et al. 2012). However, due to the prevalence of IT use, the alignment between IT and business strategies has become so institutionalized that it is no longer a primary source of differentiation in firm performance (Palmer and Markus 2000). Thus, advancing knowledge on strategic IT alignment requires understanding how IT contributes to business strategy and consolidates overall value co-creation (Coltman et al. 2015) instead of merely inferring IT's effectiveness and contributions from its fit with business strategies (Chan 1997). In addition, current researchers more strongly encourage firms to deploy tangible artifacts, rather than abstract entities, to fit with their business strategies (Coltman et al. 2015; Preston 2014).To this end, both practical contributions and a greater theoretical understanding of strategic IT alignment can result from focusing on social media—a universal IT artifact available to nearly every firm—and strategically exploiting UGC to align IT with business strategy.

Based on an extended framework of the strategic alignment model (Maes 1999), in which information on IT and business strategies are synchronized, we aimed to reveal how a firm can ensure its social media aligns with its business strategy, thereby contributing to its business performance, by effectively leveraging UGC. Specifically, in this paper, we attempted to understand how public social media opinions influence firms' strategies for innovation investment and evaluate the effectiveness of this type of investment in terms of firm performance. The resulting insight is consistent with the propositions from the prior literature, in which innovation, to sustain a firm's competitiveness, should be strategically aligned with its IT utilization (Coltman et al. 2015; Kohli and Grover 2008). To do so, we designed and developed a novel and tangible framework that can help firms delve into UGC on social media to gain valuable knowledge and validated the ways in which the knowledge extracted from our designed approach contributes to firm performance in terms of rationally strategic decisions.

Operationally, for a sample of 886 publicly-listed Chinese firms, we collected multisource data sets comprising UGC from Weibo, as well as financial and operational information. We applied a series of analytical methods, including a novel framework for sentiment analysis and econometric analysis, to reveal how public social media opinions influence firms' decisions regarding innovation investment. The primary reason for our focus on listed firms is that such firms are closely tied to the users who contribute to social media UGC. In particular, compared to those of private firms, the operations and performances of listed firms hinge much more heavily on public voices because such opinions significantly influence individual investors' and consumers' attitudes toward the firms (Kim and Youm 2017). In this regard, listed firms are keener than private firms to effectively manage UGC on social media, as doing so can help them enact relevant strategies and improve their performance.

Aside from this work's contributions to the theoretical consolidation of strategic IT alignment and to the framework of sentiment analysis, it provides three important additions to the existing literature. First, we proposed an advanced sentiment-analysis method that reveals a curvilinear (i.e., U-shaped) relationship between public-opinion valence and innovation investment. This finding reconciles the arguments from prior studies about whether positively or negatively framed information is more valuable (Chevalier and Mayzlin 2006; Sparks and Browning 2011). Second, we found the volume of public opinions to moderate the aforementioned curvilinear relationship. In particular, our findings contradict the common notion that more popular opinions have greater effectiveness when valence is equal (Flanagin and Metzger, 2013). Third, we reinforced the positive association between innovation investment and firm performance in terms of financial indicators (Belderbos et al., 2004; Grabowski and Vernon, 1990). In sum, UGC embedded in social media may not directly contribute to a firm's performance, but it does serve as a stepping stone through which firms can strategically align their resources to achieve a competitive advantage.

The rest of this manuscript is organized as follows. In Section 2, we present the theoretical development of this research, including the overarching theory (strategic IT alignment) and a comprehensive review of effective social media use; in particular, we discuss social media's implications for innovation in detail. In Section 3, we present our research hypotheses. In Section 4, we present the design of the sentiment-analysis framework, research methods, data analysis, and results. In Section 5, we discuss this work's implications and limitations, and conclude the study.

2. THEORETICAL DEVELOPMENT

2.1. Strategic IT Alignment

The primary research interest in prior studies of strategic IT alignment was understanding how firms can effectively deploy IT to strengthen their business performance (Bergeron et al. 2004). Researchers have conducted many studies to empirically unveil the antecedents and consequences of IT-business alignment. In these works, researchers have identified various antecedents of alignment, including environmental uncertainty (Sabherwal and Kirs 1994), enterprise-architecture maturity (Bradley et al. 2012), executives' IT knowledge (Hussin et al. 2002), the value-chain information intensity (Kearns and Lederer 2003), and prior successful IT implementation (Chan et al. 2006). For instance, Hussin et al. (2002) studied 256 small- and medium-sized enterprises in the U.K. and showed that a firm's level of IT mutuality and its CEO's commitment to IT significantly influenced the fusion of business and IT strategies; Wu et al. (2015) surveyed 131 large firms in Taiwan and validated the idea that IT governance helps harmonize IT-strategy alignment.

Although these studies included considerable contributions that expanded the theoretical boundaries and enriched the applications of the strategic IT-alignment paradigm, they had several weaknesses. As suggested by Luftman et al. (2017), previous researchers have been prone to conceptualizing IT-business alignment as a static state instead of an evolving process. In other words, the alignment between IT and business strategies has been examined in an ad hoc manner with abstract scales. In addition to causing potentially biased results, this method does not provide adequate value for practitioners. Additionally, prior scholars mainly developed their studies by building upon the strategic-alignment framework (Henderson and Venkatraman 1989), in which the alignment between IT and business comprises processes, structures, and people. In other words, these scholars attempted to understand how information systems can be adapted to fit with business strategies. However, the prevalence of IT use means that IT-business alignment has been largely institutionalized, so strategic IT alignment may not serve as a primary antecedent for firms' outperformance of their rivals (Palmer and Markus 2000). In this regard, instead of discussing whether IT improves business strategy, it is more crucial to understand how IT helps firms to develop business strategies that enable sustainable competitiveness.

To delve into this research direction, we employed an extension of the strategic IT-alignment framework (Maes 1999) in which IT-business alignment is synchronized through information. This approach echoes those of prior scholars who accentuated that the significance of IT's value should not be limited to enterprise systems and should include the system's information and data (Goes 2014). In addition, scholars (e.g., Coltman et al. 2015; Drucker 2015) have suggested shifting the investigation of IT alignment away from specific organizations and toward value-co-creation networks that involve many stakeholders. IT-business alignment is ideally positioned to support firms in achieving sustainable competitiveness by improving their innovation performance (Grover and Kohli 2012).

To this end, we attempted to understand how firms exploit social media UGC to support innovation and business performance, leading to contributions that complement the existing literature on strategic IT alignment in two ways. First, instead of offering an ad hoc evaluation of IT-business alignment, we developed a tangible system for applying sentiment analysis and extracting knowledge from UGC in social media, thereby validating the role that such extracted knowledge plays in supporting firms' strategic decisions regarding innovation investment and performance. Second, our study expanded the theoretical boundary of strategic IT alignment by shifting away from an investigation of a direct IT-business linkage and toward an examination of how externally sourced data facilitate IT-business synergy.

2.2. Using Social Media to Add Business Value

In terms of business practices, social media's primary value is its tremendous amounts of UGC. Thus, by investigating UGC in social media, public firms can gain insight directly from the market (i.e., from their own customers, their competitors, or even their competitors' customers) and use that information to improve their business performance (He et al. 2013). However, transforming the UGC data set into a business asset is highly contingent upon the research context; this has resulted in varied metrics and implementations in prior studies. Here, UCG or social media refers to contents shared in social media regarding the practice of a business. In our research context, stakeholders of firms share their opinions towards the products and service of firms in social media platform. Generally, researchers have taken three perspectives on effective social media use for business.

The first perspective relates to design-science research, which relates to the development and implementation of methods or tools for analyzing UGC from social media (Aral et al. 2013). IT artifacts, including cutting-edge technologies such as text mining and natural language processing, enable researchers and practitioners to better observe and understand customers' opinions (Mostafa 2013) and emotions (Thelwall et al. 2010), as well as market trends (Khadjeh Nassirtoussi et al., 2014)—all with the support of summative and visual reports. Although such descriptive output allows users to intuitively interpret public opinions, neither the underlying mechanism connecting UGC and business performance nor this mechanism's implications are well understood due to a lack of understanding of the relationship's causality.

The second perspective concentrates on business logic related to social media, such as social media's role in reshaping conventional businesses (Paniagua and Sapena 2014; Yan et al. 2016) and the business value of customer opinions or engagement (Mudambi and Schuff 2010). Although the above studies provided insightful theoretical implications, the measurements used in these studies, such as the number of words of UGC, do not enable researchers to remove noise or biased information from large-scale UGC. In other words, the methods used in the previous studies were not sophisticated enough to completely resolve the business problem.

The last perspective in this domain can be regarded as intermediate between the two aforementioned approaches; in this perspective, fundamental analytical methods are applied to Internet-driven domains such as e-commerce (Berezina et al. 2016), multichannel marketing (Goh et al. 2013), and crowdfunding (Kang et al. 2017). In sum, these studies apply to a particular organization or value chain in which business can be seamlessly integrated with social media. The findings of such case-based studies may be of limited use to firms' distinct business models, especially those whose business does not heavily rely on Internet services (e.g., manufacturing or industrial firms). To this end, it is imperative to propose a relatively generalizable approach to better understand how business value can be extracted from social media.

The UGC in social media typically comprises freeform or unstructured data (Liu et al. 2016) that cannot be effectively transformed into valuable knowledge using conventional analytical approaches such as a simple word count or quantification of consumer-review ratings (Chevalier and Mayzlin 2006; Mudambi and Schuff 2010). These methods cannot be used to extract core values (i.e., customer opinions) from massive data sets. Therefore, sentiment analysis, which evolved from text mining and natural language processing, has been deployed to understand UGC in greater depth.

After comprehensively surveying the prior literature, we noted three common sentiment-analysis approaches based on lexicons, linguistic analysis, and machine learning. In the lexicon-based approach, string matching is used to count both positive and negative indicators (terms), as listed in predefined sentiment lexicons (Pang and Lee 2008). This approach has two notable advantages: ease of implementation and speed. Due to its intuitive outputs (e.g., a large number of positive terms indicate a strong positive orientation), the lexicon-based approach has been widely adopted in business practice and research (Liu et al. 2010). However, this method suffers from low recall (the rate of detected targets out of all targets for detection) because it strongly relies on the completeness of the sentiment lexicons. Furthermore, this approach may lead to misleading outputs because of synonymy and polysemy issues.

In a machine learning-based approach, a set of labeled training data is used to train a classifier so that it can learn certain rules (Witten and Frank 2005). Afterward, these trained classifiers are used to make predictions for unlabeled data based on the rules they previously learned (Pak and Paroubek 2010; Pang et al. 2002). This process is generally carried on the text-fragment level (e.g., a sentence or document) instead of on the word level. However, the overall rules and prediction processes are a black box for users, which results in a dilemma when the explanation or improvement of an algorithm is needed.

Researchers have used the linguistic approach to attempt to understand the semantic meaning of text and to draw conclusions based on this meaning (Wilson et al. 2005); this process is similar to that of human cognition. However, due to the complexity and flexibility of human language (e.g., negation and idioms), this approach cannot be easily implemented in practice.

Our work thus incorporates the most widely used lexicon-based approach; we also adopted an automatic lexicon-expansion method to improve the completeness of the lexicon. Moreover, our method involves constructing and updating the lexicon based on domain knowledge, thereby drastically alleviating the potential bias due to synonymy and polysemy issues.

In terms of applications, sentiment analysis is also prevalently used in information systems studies to resolve business or societal questions. For instance, Mostafa (2013) employed an expert-predefined lexicon to create an application using sentiment analysis and then used that application to investigate hidden patterns in consumers' opinions toward international brands. Stieglitz and Dang-Xuan (2013) used a commercial software package, SentiStrength (sentistrength.wlv.ac.uk), to analyze the level of sentiments in politically relevant UGC from Twitter. These authors found that emotionally charged information was correlated with both the frequency and the velocity of the online transmission. Goh et al. (2013) applied sentiment analysis to investigate both marketer-generated content and UGC on a brand's Facebook fan page community, finding that UGC valence had a stronger influence than marketer-generated content valence on consumer-purchasing behaviors. These findings provide sound evidence that sentiment analysis is a reliable and effective tool for investigating how best to exploit the opinions from UGC in the decision-making process.

Remarkably, in addition to this contribution to online marketing and community building, UGC from social media serves as an important supplement to product and service innovations. In other words, firms can leverage social media UGC to collect and internalize relevant external information. This operation echoes the paradigm of inbound open innovation, in which companies use external stakeholders (e.g., customers, suppliers, or competitors) to gain valuable information for improving their internal innovation performance (Sidhu et al. 2007) or for exploring potential business opportunities (Foss et al. 2013). Scholars have discussed how to cultivate internal innovation through external information, with a particular focus on knowledge from a large user base (Chatterji and Fabrizio 2014; Garriga et al. 2013; Laursen and Salter 2006). Users' social media opinions regarding products or services can be transformed into more fine-grained and comprehensive knowledge with the help of text-mining techniques, particularly sentiment analysis. Based on a study by Liu and colleagues (2010), we used valence (the sentiment of that content) and volume (the amount of communication) as the two primary indicators for how UGC influences strategic decisions regarding innovation investment. We present the details of the development of our hypotheses in the next section.

3. HYPOTHESIS DEVELOPMENT

UGC from social media serves as important IT resources helping firms enact business strategies. In this section, we delve into the prior research on the business value of UGC and deduce the interplay between UGC and business strategy. In particular, we focus on valence and volume as the two primary dimensions used to depict the dynamics of social media in prior studies (Goh et al. 2013; Liu et al. 2010; Qiu et al. 2012) and reveal how such IT resources can be instrumental in enhancing business value. Contextually, we attempt to understand how public opinions regarding firms' products or services, as manifested in social media UGC, influence these firms' innovation-investment decisions and, as a result, their overall performance.

3.1. Valence of Opinions in UGC

Valence refers to the overall orientation (positive or negative) and intensity (weak or strong) of sentiments expressed within texts (Lerner and Keltner 2000). To date, scholars have not reached a consensus on whether negative or positive UGC opinions exert a greater influence on decision making (Derks et al. 2008; Liu et al. 2010; Yin et al. 2014b). After reviewing the literature for both sides, we inferred the ways in which firms align their business strategies with communal IT resources.

Those who have argued that negative emotions are the most influential in decision-making note that negative stimuli tend to elicit stronger and faster reactions than neutral or positive stimuli (Baumeister et al. 2001), and negative stimuli are also the most prevalent type in studies with online settings. For instance, compared to either neutral or positive information, negative information was found to have a greater influence on people's behavior (Yin et al. 2014b); in social media specifically, negative information was found to be more liable to trigger arousal and curiosity and to facilitate information diffusion (Kimmel and Kitchen 2014). In our research, we presume that social media serves as an alternative source of information that helps firms strategically align their innovation resources. Thus, compared with information that has a neutral or positive slant, negative content is more easily noted.

Negative information is thus assigned a higher weight than positive or neutral information in organizational decisions, such as the alignment of innovation investment. Furthermore, researchers in marketing and information systems have argued that individuals tend to provide evaluative opinions using negative expressions (Derks et al. 2008; Liu et al. 2010; Yin et al. 2014a). In other words, for firms, when compared to positive or neutral UGC, negative UGC may contain more constructive information, such as suggestions regarding their products, strategies, or even reputations. In this regard, firms may gain more benefit in terms of product or service improvement by concentrating on negative UGC. Last but not least, researchers have also found that firms face more pressure when negative (rather than neutral or positive) UGC prevails on social media (Pedersen and Neergaard 2009). This finding is common sense; due to the rapid propagation of information on social media, negative UGC regarding certain products or services influences not only the firm's customers but also potential customers and investors (Walther et al. 2012). The resulting pressure pushes the firm to improve its products or services through innovation, thereby influencing its decisions on innovation investment. To this end, we hypothesize that public firms exposed to negative UGC are prone to invest more in innovation initiatives than those that are neutrally discussed in social media.

Although negative UGC may arouse more attention, positive UGC also exerts a large influence. For instance, positive UGC serves as an endorsement of public firms' performance. Due to positive UGC from social media, individuals, customers, and other stakeholders have increased confidence in firms' products or services (Gallaugher and Ransbotham 2010). This phenomenon echoes the concept of selective exposure proposed by media and communication researchers, who found that individuals intentionally favor information that reinforces their preexisting views over contradictory information (Pentina et al. 2018). In this regard, UGC that compliments a firm's products or services signals a positive acceptance of its products or services, which can indeed attract the firm's attention and even motivate it to invest in sustaining or even increasing its market share.

To this end, both negative and positive opinions from social media UGC should be regarded as important IT resources that ought to be better deployed to help firms make strategic decisions on innovation investment. Thus, we propose our first hypothesis:

- **Hypothesis 1 (Social Media UGC Valence and Firm Innovation Investment)**: *The valence of social media UGC with regard to firms' products or services has a significant quadratic association (U-shaped relationship) with those firms' innovation investments.*

3.2. Moderating Role of Volume

For UGC in social media, valence comprises overall sentiment, and volume measures the amount (Chen et al. 2011; Liu et al. 2010). These two dimensions have been widely used as primary indicators in the dynamics of UGC in social media (Etzion and Awad 2007; Flanagin and Metzger 2013). In this study, we argued that the volume of UGC plays a moderating role in the relation between UGC valence and innovation investment. In particular, increased UGC volume mitigates the quadratic relationship proposed in the Social Media UGC Valence and Firm Innovation Investments Hypothesis (H1).

In social media, a high volume of negative UGC may reflect market complaints or dissatisfaction regarding the focal firm or its products or services (Presi et al. 2014). Such negative UGC can cause firms to face a series of challenges, including customer churn (Malthouse et al. 2013), reputation risk (Aula 2010), and investor discouragement (He et al. 2013). Researchers have suggested that, when facing such challenging circumstances, firms tend to apply a threat-rigidity strategy, which consists of two primary reactions: restriction in information processing and constriction in control (Staw et al. 1981). In our context, the diffusion of large amounts of negative UGC regarding firms compels firms to release themselves from this dilemma rather than extracting the potential value from the negative UGC. In the meantime, such firms ought to apply constriction in control by suspending regular operations and endeavoring to resolve the crisis as best they can. Both these reactions result in firms centralizing their resources to resolve challenges rather than undertaking proactive but risky actions such as innovation investment (D'aunno and Sutton 1992). Thus, firms are discouraged from investing in innovation due to the dramatically increasing volume of negative UGC from social media.

However, considerable amounts of positive UGC may also deter subsequent innovation investment. When receiving a high volume of positive UGC, firms may not be eager to make significant innovation investments, such as by transforming stakeholders' external knowledge into to internal innovation initiatives; this is due to organizational inertia, in which firms tend to continue on their current trajectories (Kelly and Amburgey 1991). Contextually, positive UGC signals that the public is praising the focal firm's current performance (He et al. 2013). Firms thus see the wisest course of action as keeping the current strategy or making a slight improvement or adjustment, in lieu of taking risks by investing in disruptive innovations. Therefore, firms resist drastic investments in innovation if their products or services are receiving a considerable volume of positive UGC from social media.

Thus, we propose the following counterintuitive hypothesis:

- **Hypothesis 2 (Social Media UGC Moderation):** *Increased social media UGC to firms mitigates the quadratic association between the valence of that UGC and the firms' innovation investments.*

3.3. The Consequences of Innovation Investment

Researchers have clearly demonstrated the positive role that innovation investment plays in improving financial performance (Belderbos et al. 2004; Grabowski and Vernon 1990; Levin 1988; Sougiannis 1994). Innovation is a key part of firms' ability to compete with rivals and sustain their competitiveness (Cooke and Wills 1999; Morgan and Berthon 2008). In a fiercely competitive market, breakthrough products or services can help firms attract more customers (Katila and Ahuja 2002) and, consequently, outperform their rivals.

Scholars have argued that a reliance on externalities (i.e., external information or stakeholders) is conducive to innovation performance and that it thus eventually contributes to business performance (Chesbrough et al. 2006). Accessing UGC from social media can help firms understand marketing trends and public voices but exploiting such external knowledge in products or services requires a tangible innovation investment. In other words, the ideas, inspirations, and suggestions from UGC can be effectively incorporated into product or service innovation only if a firm is committed to investment in innovation (Cohen and Levinthal 1990; Robertson et al. 2012), from which the firms can gain tangible benefits in terms of overall performance (Berchicci 2013). Scholars have also found innovation investment to help reduce operational costs (Aw et al. 2011) and increase operational efficiency (Aw et al. 2008; DiMasi et al. 2003; George et al. 2002), both of which contribute to improved firm performance. In this regard, it is reasonable to assume that there is a positive correlation between a firm's innovation investment and its business performance. Therefore, we propose the following hypothesis:

- **Hypothesis 3 (Firm Innovation and Performance):** *There is a positive association between the extent of a firm's innovation investment and that firm's performance.*

4. RESEARCH METHOD

4.1. Sampling and Data Collection

To test our theoretical hypotheses, we examined the impact that social media opinions had on firm innovation by collecting data from A-share firms listed on the Chinese stock markets in the Shanghai and Shenzhen Stock Exchanges. As mentioned previously, alignment between social media UGC and business strategy is more instrumental in helping publicly listed firms (compared to private firms) improve their business performance. There are two primary reasons. First, investors can observe any fluctuations in public firms' UGC. On the one hand, an increase in positive opinions can build these investors' confidence. On the other hand, investors may become unnerved by the emergence of negative or critical UGC toward the focal firms, resulting in financial losses (Kim and Youm 2017). Second, the majority of A-shares are issued by state-owned enterprises in China, which have struggled to reform and thus align with the transformation of China's economy (Jefferson 2016). To achieve the goal of reformation, innovation in product or service is crucial. Thus, we provided relevant ways in which listed firms can leverage social media UGC to outperform rivals in both innovation and business practices. Nonetheless, the exclusion of private firms and small and medium-sized enterprises in our sample is a potential limitation in our study's implications, as highlighted in the subsequent section.

We constructed our sample by collecting data from multiple sources through a series of operations. First, we collected firms' demographics and performance information from both Compustat and GTA (www.gtafe.com). We used Compustat to obtain firms' financial indicators and GTA to obtain information about firms' innovation investment and demographics. We selected 886 firms that were listed in both databases and free of missing values in the studied variables to construct our strictly balanced panel; the resulting data set included 4430 observations, one per year per company from 2011–2015. Second, we developed a web crawler with the granted API to collect these companies' UGC from Weibo. Weibo is one of the largest and most influential social media platforms in China. We used the firms' names as our keywords for data crawling. We adopted a name-entity recognition method (Stanford CoreNLP, stanfordnlp.github.io/CoreNLP) to recognize the products and services included in the public posts. It is well known that content is social media is noisy. To minimize irrelevant information and confirm the method's accuracy, we conduct a semi-heuristic method to filter crude data to obtain an innovation-relevant (related to products or services) social media corpus. We first manually construct an innovation indicator lexicon (e.g., design, invent, development, etc.). Then, we search for widely used synonyms or alternatives in social media to expand these indicators. Finally, we filter these social media content using these indicators. Next, because UGC comprises unstructured and freeform data with numerous abbreviations, we apply abbreviation-detection methods (Kiss and Strunk 2006; Malviya et al. 2016) to identify and construct an abbreviation set for each firm. Independent assistants confirmed the accuracy of the abbreviation sets. In all, we collected 6.2 million microblog posts in our sample.

4.2. Novel Framework for Sentiment Analysis

Sentiment embedded in a text[1](e.g., a document or microblog) can be mined by means of sentiment-analysis techniques (Pang and Lee 2008). The mostly widely used method involves extracting sentiment indicators (e.g., positive and negative terms) based on a predefined sentiment lexicon and then evaluating the overall sentiment based on a statistical analysis of these indicators. However, there are two challenges to this approach. First, it strongly relies on the completeness of the sentiment lexicon. To guarantee the completeness of the sentiment lexicon, at least two measures need to be taken: The lexicon must be expanded as much as possible to cover various domains, and it must be updated in a timely manner to keep up to date with the data stream. However, these measures were proposed for ideal settings,

[1] There are two ways to analyze UGC and to calculate the yearly valence for each firm. The first is to extract the sentiment of each post and then calculate the overall sentiment according to the sum of the individual sentiments. The second is to combine all the posts into a single document and then extract the sentiment embedded in that document. Given that social media posts are usually very short and that statistical methods usually do not perform very well on short texts, the second approach is widely used for text mining and sentiment analysis of social media posts. We also adopted topic-modeling methods as part of our domain-sentiment construction process; this was very important to the effectiveness of our final sentiment analysis. As scholars have proven that topic-modeling methods suffer an obvious performance reduction when applied to short text, we applied one of the most widely used methods for dealing with this problem: combining short texts with similar characteristics into one long text (Hong and Davison 2010). We used this method, which is the second sentiment-analysis approach described above.

and they are difficult to implement in real cases. Second, the output accuracy may be biased due to varied contextual settings. In sentiment analysis, the polarity of a target term may vary based on the context or business domain. For instance, the adjective “heavy” is mostly interpreted as positive when applied to an investment but may be negative when describing an electronic product such as a laptop or smartphone. This situation is prevalent in Chinese data mining because Chinese (as is well-known) is quite vague and has strong elasticity (Su et al. 2008; Zhang et al. 2009).

To overcome this challenge, we adopted a topic-modeling method to expand the existing sentiment lexicon and construct timely, domain-specific sentiment lexicons. In a topic-modeling process, such as *latent Dirichlet allocation* (LDA), each text segment (UGC from a microblog post in our context) is assumed to be characterized by a mixture of latent topics, each of which can be characterized by a mixture of terms. We took two steps to construct each text segment (e.g., document). First, we sampled a topic according to the document-topic distribution. Second, we sampled a word according to the topic-word distribution. In this way, we revealed the characteristics of the given document through its topic distribution, and we described the topic’s characteristics using its word distribution. In practice, individuals tend to employ terms with similar meanings as alternative expressions. In other words, there is a higher probability of co-occurring terms (words) with similar meanings in a short text fragment (e.g., a sentence or a window) than in a longer fragment. During the sampling process, a topic-modeling method (e.g., LDA) is capable of determining the probability of a co-occurrence in a certain context and of clustering terms (based on probability of co-occurrence) into relevant topics accordingly. We used this method to cluster terms (words) with similar meanings in a certain context by topic.

However, despite the aforementioned improvements, we faced two key challenges when implementing LDA in practice. First, determining the optimal number of topics was challenging. An improper topic number could have biased the topic distribution and thus destabilized the term cluster. To overcome this challenge, we adopted a perplexity-based approach to select the optimal number of latent topics for a given corpus (Steyvers et al. 2004). More specifically, this perplexity score served as a proxy indicator for the number of topics (i.e., a lower perplexity scores indicated a higher optimality). After a series of testing rounds, we selected the perplexity score with the lowest value. In particular, we defined the perplexity score of a corpus D as follows:

$$perp(D) = \exp\left[-\frac{\sum_{d\in D} lnPr(d|\theta,\varphi)}{\sum_{d\in D}|d|}\right],$$

where d is a text segment (e.g., a microblog post) that comprises a set of words. The conditional probability $Pr(d|\theta,\varphi)$ is the generation probability of all the terms $w \in d$.

Another challenge resulted from the identification of a label for each topic. We struggled to identify and select positive and negative topics. We employed an original sentiment lexicon as the seed. While treating each topic as a bag of words, we calculated the topic's distance from both the original positive and original negative lexicons using similarity analysis. In this step, we used the probability of each word within a topic as a weight to describe its importance in composing that topic. Then, we determined the top-ranked topics (e.g., topγ) for both the original positive and original negative lexicons. Given that there was noise in the words comprising each topic, we executed a supplementary procedure. For the selected topics, we then conducted another round of ranking and selection; this round was based on the importance of each word in the selected topic (e.g., topε). In other words, the newly constructed domain sentiment lexicon contained only the topics with refined words. We deployed the Chinese sentiment analysis lexicon HowNet (Hong and Davison 2010) as the original sentiment lexicon for this work. The design of our sentiment analysis and the entire workflow process are illustrated in Figure 1. More information about Chinese and English text mining can be found in the Appendix.

INSERT FIGURE 1 HERE

After constructing the domain's sentiment lexicon, we extracted the positive and negative words for that domain from the UGC relating to each firm. Notably, instead of extracting the sentiment from each piece of UGC, we combined the UGC for each firm into a single document and extracted the sentiment from those documents. This operation is widely used in text mining and sentiment analysis for short texts such as microblog posts and comments due to the limitations of statistical methods when applied to such short texts (Hong and Davison 2010). To determine public opinion for each firm, we used a sentiment score (a relative value) to summarize people's orientation toward the firm's products or services in a certain period; we did this instead of simply counting the corresponding number of positive and negative words. The following is the mathematical expression of the sentiment score:

$$dm\_sent = \frac{dm\_pos - dm\_neg}{dm\_pos + dm\_neg},$$

where dm_pos and dm_neg are the number of positive and negative words, respectively, from the domain. The value of dm_sent ranges from -1 to +1, where -1 indicates that no positive content was posted in a certain period and +1 indicates that no negative content was posted in that period. The null value (i.e., 0) indicates there were as many positive as negative contents posted; the UGC in this case can be regarded as neutral.

4.3. Measurement

4.3.1. Key Variables

Using the analytical approach proposed 2, we measured the valence of a firm's UGC from Weibo in a certain financial year and denoted this as $SENT_{it}$. To measure the volume of this UGC, we counted the number of Weibo posts containing information related to the focal firm within the given financial year and denoted that as VOL_{it}.

We adopted each firm's yearly research and development investment as a proxy for innovation investment; scholars have widely adopted this proxy to measure the extent to which firms invest in innovation (e.g., Czarnitzki and Hottenrott 2011). We measured firm performance as its annual revenue. Intuitively, firms that generate more revenue can be thought of as having better performance. We denoted these two variables, which we obtained from both the GTA and Compustat databases (with cross-validation), as RDI_{it} (innovation investment) and RVN_{it} (revenue).

4.3.2. Control Variables

We also included a rich set of control variables, both to explore other potentially influential factors in terms of innovation investment and firm performance and to limit the threat of endogeneity (Wooldridge 2010).

Number of researchers ($RCHR_{it}$): Innovation is a process of knowledge recombination and creation. Researchers can link existing and new knowledge to transform ideas or innovation knowledge into products or services. Moreover, research expenses (e.g., salaries, equipment costs, and recruitment fees) account for part of research and development spending. Therefore, it is imperative to include these expenses in the model as a control variable.

Profit (PTF_{it}): Innovation is an important approach for gaining competitive advantage and increasing profits. From this point of view, both low-profit and high-profit firms will invest in innovation to either improve or sustain their business. Thus, we controlled for each firm's profit as well.

Asset (AST_{it}): Innovation is risky, as it may result in failure. In this regard, firms with more assets have a higher capability of tolerating and managing these risks. Therefore, we included each firm's assets as a factor that could affect that firm's innovation investment.

Return on assets (ROA_{it}): In organizational studies, this variable is widely adopted to evaluate how well a firm operated during a financial year. Firms with high return on assets have higher investment capacities because they have more resources. Therefore, we controlled for return on assets in our research model.

Employee number (EMP_{it}): We used the number of employees to describe the size of each firm. Intuitively, investment spending varies for firms of different sizes. Therefore, we included firm size as a control variable in our model.

Industrial categories (IND_i): Preferences regarding innovation investment vary across industries. Firms in high-technology domains such as biotechnology, pharmacology, and IT are likely to spend more on innovation than other firms are. Therefore, we constructed a categorical variable, IND_i, to control for variations in innovation investment across industries.

Year dummy ($YEAR_t$): Finally, Weibo's advancement could change social media's impact on firms over time. Thus, we included year dummy variables in our estimation models.

The definitions and the descriptive statistics for all of the studied variables are summarized in Tables 1 and 2. To mitigate the possibility of reverse causality, we lagged all independent and control variables by one year in our analyses.

INSERT TABLE 1, TABLE 2 HERE

4.4. Estimation Procedure

Table 3 reports the correlations among the studied variables; the majority of the bivariate correlations were below the recommended 0.70 threshold level. To evaluate the threat of collinearity, we calculated the variance inflation factors for all of the regression models. The maximum estimated variation inflation factor was 6.21, which is lower than the recommended ceiling of 10.0 (Cohen et al. 2003). In an attempt to address the issue of endogeneity, we included an extensive set of control variables and lagged dependent variables in the estimation models.

INSERT TABLE 3 HERE

To validate our hypotheses, we developed empirical models to assess the explanatory power of each variable. As noted previously, we constructed a set of strictly balanced panel data for 886 firms collected from 2011 through 2015. We first examined the influence of social media UGC on innovation investment (i.e., the Social Media UGC Valence and Firm Innovation Investment Hypothesis, H1 and the Social Media UGC Moderation Hypothesis, H2). Specifically, our models estimated the impact of valence ($SENT_{it-1}$) on future research and development investment (RDI_{it}); we also tested the moderating effect of online-post volume. We also examined the impact of innovation investment on firm performance (i.e., the Firm Innovation and Performance Hypothesis, H3). Our empirical models can be described as follows:

$$\begin{aligned}\log(RDI_{it}) = {} & \beta_0 + \beta_1 \mathrm{SENT}_{it-1}^2 + \beta_2 SENT_{it-1} + \beta_3 \log(VOL_{it-1}) * \mathrm{SENT}_{it-1}^2 + \beta_4 \log(VOL_{it-1}) \\ & + \beta_5 \log(VOL_{it-1}) * SENT_{it-1} + \beta_6 \log(RCHR_{it-1}) + \beta_7 \log(PFT_{it-1}) \\ & + \beta_8 \log(AST_{it-1}) + \beta_9 ROA_{it-1} + \beta_{10} \log(EMP_{it-1}) + \beta_{11} IND_i + \beta_{12} YEAR_t + \varepsilon_{it-1}\end{aligned}$$

$$\begin{aligned}\log(RVN_{it}) = {} & \gamma_0 + \gamma_1 RDI_{it-1} + \gamma_2 \log(RCHR_{it-1}) + \gamma_3 \log(PFT_{it-1}) + \gamma_4 \log(AST_{it-1}) + \gamma_5 ROA_{it-1} \\ & + \gamma_6 \log(EMP_{it-1}) + \gamma_7 IND_i + \gamma_8 YEAR_t + \xi_{it-1}\end{aligned}$$

In the above specifications, subscripts denote measures across firms (i) and years (t). Due to the variation in the firms' sizes in the sample, we transformed all firm-size-related variables, including the dependent variables (RDI_{it} and RVN_{it}), using logarithms.

The results are presented in Table 4. Specifically, we regressed innovation investment on the squared valence term and on the interaction term of volume and squared valence. We also included other terms, such as valence and the interaction between valence and volume, in the models, as suggested by Wooldridge (2010). We applied both random-effects and fixed-effects models. In Model 1 and Model 3, we included only the key predictors, using random and fixed effects, respectively; in Model 2 and Model

4, we also presented the results from the full model, taking into account random and fixed effects, respectively.

To validate the Social Media UGC Valence and Firm Innovation Investment Hypothesis (H1), we examined the influence of squared valence on innovation investment in both Model 2 (random effects) and Model 4 (fixed effects). The results revealed that squared valence had a consistent positive correlation with innovation investment ($b = 4.158$, $p < 0.01$ in Model 2; $b = 4.823$, $p < 0.01$ in Model 4). These significant results revealed that valence had a U-shaped relationship with innovation investment, strongly supporting H1.

The Social Media UGC Moderation Hypothesis (H2) related to the moderating role of volume. Model 2 and Model 4 both indicated that the interaction term between volume and squared valence had a negative relationship with innovation investment ($b = -0.767$, $p < 0.01$ in Model 2; $b = -0.913$, $p < 0.01$ in Model 4). This finding indicates that the relation between squared valence and innovation investment is reduced as volume increases. This result strongly supports H2.

INSERT TABLE 4 HERE

We examined Firm Innovation and Performance Hypothesis (H3) by regressing revenue on innovation investment. The results are reported in Table 5. From the estimation results, we found that innovation investment was positively associated with revenue in the following year for both the random-effects model ($b = 0.049$, $p < 0.01$ in Model 5) and the fixed-effects model ($b = 0.004$, $p < 0.01$ in Model 6). These results are consistent with the findings noted in the previous research (Belderbos et al., 2004; Grabowski and Vernon, 1990; Levin, 1988; Sougiannis, 1994), which highlighted the importance of investing in innovation to garner advantages for future business performance.

INSERT TABLE 5 HERE

4.5. Robustness Checks

We included several post hoc analyses to assess the robustness of our findings. For instance, we examined the proposed hypotheses by extracting two samples (based on firm size) from the data set. We selected the top 20% and bottom 20% of firms by size (number of employees) and used these groups to re-estimate our hypotheses. The estimation results for the top 20% of firms by size are listed in Table 6. Based on the data listed in Table 6, the Social Media UGC Valence and Firm Innovation Investment Hypothesis 1 (H1) and the Social Media UGC Moderation Hypothesis (H2) were supported for the largest firms.

INSERT TABLE 6 HERE

We also carried out the same test on the bottom 20% of firms by size. The estimation results for the smallest firms are listed in Table 7 below. These results also confirm H1 and H2. We also used the models to evaluate the impact of innovation investment on revenue. These results were consistent with our findings from the prior analyses. All of our results suggest that our empirical analyses are robust.

INSERT TABLE 7 HERE

5. DISCUSSION AND CONCLUSION

In this study, we revealed the U-shaped relationship between a firm's public-opinion valence and its innovation investment. This finding provides evidence that public opinions expressed via social media influence firms' strategic decisions pertaining to innovation investment. More specifically, strongly negative opinions (the left part of the U shape) are the most likely to cause firms to invest more in innovation, as they seek to improve the undesirable status quo; firms may also be motivated to invest more in innovation when they receive praise.

Interestingly, we also found that the volume of social media UGC mitigated the curvilinear relationship between a firm's UGC valence and its innovation investment. Specifically, a sharp curve indicated that a firm promptly and drastically responded to public opinion; in other words, a small change in the independent variable led to a large change in the dependent variable. A flatter shape meant that the dependent variable was less sensitive to the changes in the independent variable. These findings indicate that, on social media, firms respond more slowly to opinions from crowds when the volume of such opinions is higher. This situation is easy to understand in a real-world setting. Firms that receive a large volume of negative feedback consider allocating resources to resolve the immediate dilemma (e.g., customer churn or reputation risk) more desirable than increasing innovation investment. Firms that receive large amounts of praise or acknowledgement see maintaining the status quo as more realistic than making huge investments that could create disruptive actions. Firms are supposed to adjust their innovation-investment strategies to fit with the fluctuations of social media.

Our findings from the robustness check also provide clear evidence in the controversial debate regarding whether the volume of UGC (number of posts) alone is adequate for depicting a social media trend (Chevalier and Mayzlin 2006; Mudambi and Schuff 2010). Our results reveal that this method is less precise, especially for particularly large or small firms (by number of employees). This corroborates the significance of leveraging advanced techniques to analyze data that may contain considerable biased or noisy information—such as social media UGC. Thus, our proposed method complements traditional measurement in characterizing the dynamics of social media using the big data approach.

5.1. Theoretical Contributions

This study provides four primary contributions to the existing theory and literature. First, this work expanded the theoretical boundary of strategic IT alignment. Although Luftman et al. (2017) extended the framework of strategic alignment to highlight information's role in creating synergy in the alignment of

IT and business strategies, the underlying mechanism (i.e., how information can be managed to achieve IT-business fusion) was not previously fully articulated. In this work, we contextualized this theoretical paradigm to describe how firms can exploit information from social media UGC to make better strategic decisions (regarding innovation investment) and to thus improve their overall performance. In particular, we developed a tangible framework for extracting business value from digital artifacts. This attempt was a response to previous studies in which researchers criticized the examination of IT-business alignment using ad hoc or abstract scales (Luftman et al. 2017).

Second, we proposed and validated a novel framework of sentiment analysis. In particular, we adopted a topic-modeling method to extend the existing sentiment lexicon to construct time-domain-specific sentiment lexicons. This new method enabled us to more accurately capture UGC valence from social media and to mitigate bias from traditional sentiment analysis, which is particularly helpful in Chinese data mining. Notably, our work extended prior studies that accentuated the technological aspects of sentiment analysis (Khadjeh Nassirtoussi et al. 2014; Mostafa 2013; Thelwall et al. 2010) to highlight the application of sentiment analysis in resolving business problems. In other words, we provided a solution for transforming data sets into business assets.

Third, our work expanded the theoretical boundary of the impact of social media on the strategic alignment of innovation investment. In previous studies, researchers widely discussed the value of exploiting UGC in business practices in terms of downstream activities, such as marketing campaigns (Goh et al. 2013), individual investment behaviors (Chen et al. 2014), and the design of recommendation systems (Zhou et al. 2012). However, few scholars have investigated how to leverage UGC from social media to support upstream activities (i.e., strategic decisions regarding innovation investment) in our context. Our findings imply that UGC from social media can also be used to support firms' strategic decisions. In particular, our findings will be instrumental in helping firms whose business does not heavily rely on the Internet to leverage Internet-oriented resources and thus improve their strategic decisions and overall performance.

Last but not least, our work contributed to the social media literature by describing the quadratic relationship between social media UGC valence and innovation investment. Although the importance of UGC valence was widely recognized in prior studies (Pentina et al. 2018; Qiu et al. 2012; Rim and Song 2016), the question of whether positive or negative content has a larger influence on business decisions remains under debate. In this study, we outlined the curvilinear (i.e., U-shaped) relationship between valence and innovation investment. To the best of our knowledge, this is the first work to empirically reveal such a relationship. Interestingly, we found that rational voices (i.e., UGC expressing neutral opinions), when compared to irrational ones (i.e., UGC expressing positive or negative opinions), may afford less valuable information for firms to use in making strategic decisions. Our analyses also showed that volume mitigates the effect of this curvilinear relationship. This finding extends our existing understanding of the interaction between volume and valence in UGC from social media. Firms can strategically leverage communal IT resources from social media to extract knowledge and enact business strategies.

5.2. Managerial Contributions

In addition to its theoretical contributions, this work also has two important implications for practitioners.

First, we provided a clear road map for implementing a novel framework of sentiment analysis to mine public opinions. Practitioners can utilize our design to implement our proposed method for analyzing public opinions regarding focal companies. The sentiment-analysis score provides an intuitive understanding of public opinions, enabling practitioners to directly observe the trends in opinions and to use such trends as antecedents for additional analysis. In addition, Chinese culture, which is regarded as strongly collectivist, likely is more susceptible than other cultures to the word-of-mouth effect (Hong and Davison 2010). Therefore, our Chinese-based sentiment analysis is helpful for companies seeking to understand the nature of Chinese consumers.

Second, our findings provide an alternative approach for firms making innovation-investment decisions. Such investments are vital to firms' strategic alignment as they seek to preemptively collect advantages for future business opportunities. However, the questions of when and how such innovation investments should be acted upon have not been thoroughly studied. In this work, we leveraged sentiment analysis to understand public opinions embedded in social media UGC. We propose that this UGC can help firms make innovation-investment decisions. Our results suggest that firms should constantly observe public opinions and react based on the overall trends of those opinions. In particular, strongly positive and strongly negative opinions should be regarded as important signals for investment in innovation, which has been shown to increase future revenues. In addition, to avoid hazardous investments, firms should not make innovation investments purely based on UGC.

5.3. Limitations, Future Research and Conclusion

Although we demonstrated a comprehensive sentiment-analysis design and obtained interesting empirical findings, this work has some limitations. First, although researchers have widely accepted that innovation investment can increase a firm's performance, other external factors influence this impact. More control covariates—such as environmental turbulence and policy effects—should be considered in future research. Second, to improve the quality of the sentiment analysis, a more complete and precise Chinese lexicon should be created. Although the Chinese sentiment analysis lexicon we used (HowNet20, www.keenage.com) is widely accepted, it does have room for improvement, particularly regarding the refinement of adjectives' qualitative degrees. It is suggested that future work apply more sophisticated algorithms that can extract highly fine-grained samples, though we have implemented a series of systematical methods to rule out the bias in this work. Third, this study's data set includes only public firms (i.e., firms that have had initial public offerings, IPO). This sample selection helped us to investigate the impact of social media on traditional-industry firms, but the external validity of the resulting research findings is limited for other types of firms. In further research, data from other types of firms should be collected and analyzed to triangulate our empirical findings. Fourth, although we used various methods, including cross-validated topic modeling, to manage contextual meaning in opinion mining and to increase the accuracy to an acceptable level, more sophisticated methods for understanding linguistic contexts should be employed in future studies.

Although social media has been proven to foster tremendous business value, few scholars have examined how such ubiquitous IT artifacts can synergize with firms' strategies. In this study, we designed and developed a novel framework of advanced sentiment analysis. Using this powerful and reliable tool, we outlined public opinions from social media UGC and empirically validated how firms strategically aligned this IT resource with their innovation strategies. Specifically, we found a curvilinear (U-shaped) relationship between UGC valences and innovation-investment decisions, as both negative and positive UGC motivated firms to assign more resources to innovation. In addition, we found that this curvilinear relationship was mitigated when the volume of UGC increased; this finding is counterintuitive but interesting, as it suggests that firms ought to rationally manage outbursts of unipolar public opinions. Furthermore, this study provided evidence that innovation investment is conducive to outperforming one's competitors.

7. REFERENCES

Amid, A., Moalagh, M., Zare Ravasan, A. 2012. Identification and classification of ERP critical failure factors in Iranian industries Information Systems 37 (3), 227–237.

Aral, S., Dellarocas, C., Godes, D. 2013. Introduction to the Special issue—Social media and business transformation: A framework for research. Information Systems Research 24 (1), 3–13.

Aula, P. 2010. Social media, reputation risk and ambient publicity management. Strategy & Leadership 38 (6), 43–49.

Avison, D., Jones, J., Powell, P., Wilson, D. 2004. Using and validating the strategic alignment model. Journal of Strategic Information Systems 13 (3), 223–246.

Aw, B. Y., Roberts, M. J., Xu, D. Y. 2008. R&D Investments, Exporting, the evolution of firm productivity; American Economic Review 98 (2), 451–456.

Aw, B. Y., Roberts, M. J., Xu, D. Y. 2011. R&D investment, exporting, and productivity dynamics. american economic review 101 (4), 1312–1344.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., Vohs, K. D. 2001. Bad Is stronger than good. Review of General Psychology 5 (4), 323–370.

Belderbos, R., Carree, M., Lokshin, B. 2004. Cooperative R&D and firm performance. Research Policy 33 (10), 1477–1492.

Berchicci, L. 2013. Towards an open R&D system: Internal R&D investment, external knowledge acquisition and innovative performance. Research Policy 42 (1), 117–127.

Berezina, K., Bilgihan, A., Cobanoglu, C., Okumus, F. 2016. Understanding satisfied and dissatisfied hotel customers: Text mining of online hotel reviews. Journal of Hospitality Marketing and Management 25 (1), 1–24.

Bergeron, F., Raymond, L., Rivard, S. 2004. Ideal patterns of strategic alignment and business performance. Information & Management 41 (8), 1003–1020.

Bradley, R. V., Pratt, R. M. E., Byrd, T. A., Outlay, C. N., Wynn, D. E. 2012. Enterprise architecture, IT effectiveness and the mediating role of IT alignment in US hospitals. Information Systems Journal 22 (2), 97–127.

Casey, S. 2017. 2016 Nielsen social media report. Available at: www.nielsen.com/us/en/insights/reports/2017/2016-Nielsen-Social-Media-Report.html.

Chan, Y. E. 1997. Business strategic orientation, information systems strategic orientation, and strategic alignment. Information Systems Research 8 (2), 125–150.

Chan, Y. E., Sabherwal, R., Thatcher, J. B. 2006. Antecedents and outcomes of strategic is alignment: An empirical investigation. IEEE Transactions on Engineering Management 53 (1), 27–47.

Chatterji, A., Fabrizio, K. 2014. Using users: When does external knowledge enhance corporate product innovation. Strategic Management Journal 35 (10), 1427–1445.

Chen, H., De, P., Hu, Y., Hwang, B. H. 2014. Wisdom of crowds: The value of stock opinions transmitted through social media. Review of Financial Studies 27 (5), 1367–1403.

Chen, Y. L., Tang, K., Wu, C. C., Jheng, R. Y. 2014. Predicting the influence of users' posted information for eWOM advertising in social networks. Electronic Commerce Research and Applications 13 (6), 431–439.

Chen, Y., Wang, Q., Xie, J. 2011. Online social interactions: A natural experiment on word of mouth versus observational learning. Journal of Marketing Research 48 (2), 238–254.

Chesbrough, H., Vanhaverbeke, W., West, J. 2006. Open Innovation: Researching a New Paradigm. New York: Oxford University Press.

Chevalier, J. A., Mayzlin, D. 2006. The effect of word of mouth on sales: Online book reviews. Journal of Marketing Research 43 (3), 345–354.

Cohen, P., Cohen, J., West, S. G., Aiken, L. S. 2003. Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences, Abington, UK: Routledge.

Cohen, W. M., Levinthal, D. A. 1990. Absorptive capacity: A new perspective on learning and innovation. Administrative Science Quarterly 35 (1), 128–152.

Coltman, T., Tallon, P., Sharma, R., Queiroz, M. 2015. Strategic IT alignment: Twenty-five years on. Journal of Information Technology 30 (2), 91–100.

Cooke, P., Wills, D. 1999. Small firms, social capital and the enhancement of business performance through innovation programmes. Small Business Economics 13 (3), 219–234.

Czarnitzki, D., Hottenrott, H. 2011. R&D investment and financing constraints of small and medium-sized firms. Small Business Economics 36 (1), 65–83.

D'aunno, T., Sutton, R. I. 1992. The responses of drug abuse treatment organizations to financial adversity: A partial test of the threat-rigidity thesis. Journal of Management 18 (1), 117–131.

Derks, D., Fischer, A. H., Bos, A. E. R. 2008. The role of emotion in computer-mediated communication: A review. Computers in Human Behavior 24 (3), 766–785.

DiMasi, J. A., Hansen, R. W., Grabowski, H. G. 2003. The price of innovation: New estimates of drug development costs. Journal of Health Economics 22 (2), 151–185.

Drucker, P. F. 2015. Innovation and Entrepreneurship. New York: Harper Wiley.

Ernst, M., Brem, A. 2017. Social media for identifying lead users? Insights into lead users' social media habits. International Journal of Innovation and Technology Management 14 (4), 1–21.

Etzion, H., Awad, N. F. 2007. Pump up the volume? Examining the relationship between number of online reviews and sales: Is more necessarily better? In Proceedings of the 2007 International Conference on Information Systems, Atlanta, GA: AIS Press, pp. 1–14.

Flanagin, A. J., Metzger, M. J. 2013. Trusting expert versus user-generated ratings online: The role of information volume, valence, and consumer characteristics. Computers in Human Behavior 29 (4), 1626–1634.

Foss, N. J., Lyngsie, J., Zahra, S. A. 2013. The role of external knowledge sources and organizational design in the process of opportunity exploitation. Strategic Management Journal 34 (12), 1453–1471.

Gallaugher, J., Ransbotham, S. 2010. Social media and customer dialog management at Starbucks. MIS Quarterly Executive 9 (4), 197–212.

Garriga, H., Von Krogh, G., Spaeth, S. 2013. How constraints and knowledge impact open innovation. Strategic Management Journal 34 (9), 1134–1144.

George, G., Zahra, S. A., Wood, D. R. 2002. The effects of business-university alliances on innovative output and financial performance: A study of publicly traded biotechnology companies. Journal of Business Venturing 17 (6), 577–609.

Goes, P. B. 2014. Big data and IS research: Editor's comments. MIS Quarterly 38 (3), iii–viii.

Goh, K. Y., Heng, C. S., Lin, Z. 2013. Social media brand community and consumer behavior: Quantifying the relative impact of user- and marketer-generated content. Information Systems Research 24 (1), 88–107.

Grabowski, H. G., Vernon, J. 1990. A new look at the returns and risks to pharmaceutical R&D. Management Science 36 (7), 804–821.

Grover, V., Kohli, R. 2012. Co-creating IT value: New capabilities and metrics for multifirm environments. MIS Quarterly 36 (1), 225–232.

He, W., Zha, S., Li, L. 2013. Social media competitive analysis and text mining: A case study in the pizza industry. International Journal of Information Management 33 (3), 464–472.

Henderson, J., Venkatraman, N. 1989. Strategic alignment: A framework for strategic information technology management. Working paper no. 190, Center for Information System Research, Sloan School of Management, MIT, Cambridge, MA.

Hong, L., Davison, B. D. 2010. Empirical study of topic modeling in Twitter. In Proceedings of the First Workshop on Social Media Analytics, Washington, DC, pp. 80–88.

Hussin, H., King, M., Cragg, P. 2002. IT Alignment in small firms. European Journal of Information Systems 11 (2), 108–127.

Jefferson, G. 2016. State-owned enterprise in China: Reform, performance, prospects. Working paper 109R, Department of Economics and International Business School, Brandeis University, Waltham, MA.

Kang, L., Jiang, Q., Tan, C. H. 2017. Remarkable advocates: An investigation of geographic distance and social capital for crowdfunding. Information and Management 54 (3), 336–348.

Kaplan, A. M., Haenlein, M. 2010. Users of the world, unite! The challenges and opportunities of social media. Business Horizons 53 (1), 59–68.

Katila, R., Ahuja, G. 2002. Something old, something new: A longitudinal study of search behavior and new product introduction. Academy of Management Journal 45 (6), 1183–1194.

Kearns, G. S., Lederer, A. L. 2003. A resource-based view of strategic it alignment: How knowledge sharing creates competitive advantage. Decision Sciences 34 (1), 1–29.

Kelly, D., Amburgey, T. L. 1991. Organizational inertia and momentum: A dynamic model of strategic change. Academy of Management Journal 34 (3), 591–612.

Khadjeh Nassirtoussi, A., Aghabozorgi, S., Ying Wah, T., Ngo, D. C. L. 2014. Text mining for market prediction: a systematic review. Expert Systems with Applications 41 (16), 7653–7670.

Kim, E. H., Youm, Y. N. 2017. How do social media affect analyst stock recommendations? Evidence from S&P 500 electric power companies' Twitter accounts. Strategic Management Journal 38 (13), 2599–2622.

Kimmel, A. J., Kitchen, P. J. 2014. WOM and social media: Presaging future directions for research and practice. Journal of Marketing Communications 20 (1–2), 5–20.

Kiss, T., Strunk, J. 2006. Unsupervised multilingual sentence boundary detection. Computational Linguistics 32 (4), 485–525.

Kohli, R., Grover, V. 2008. Business aalue of IT: An essay on expanding research directions to keep up with the times. Journal of the Association of Information Systems 9 (1), 23–39.

Kumar, V., Rajan, B. 2012. The perils of social coupon campaigns. MIT Sloan Management Review 53 (4), 13–14.

Laursen, K., Salter, A. 2006. Open for innovation: The role of openness in explaining innovation performance among U.K. manufacturing firms. Strategic Management Journal 27 (2), 131–150.

Lerner, J. S., Keltner, D. 2000. Beyond valence: Toward a model of emotion-specific influences on judgment and choice. Cognition & Emotion 14 (4), 473–493.

Levin, R. C. 1988. Appropriability, R&D spending, and technological performance. American Economic Review 78 (2), 424-428.

Liu, X., Singh, P. V., Srinivasan, K. 2016. A structured analysis of unstructured big data by leveraging cloud computing. Marketing Science 35 (3), 363–388.

Liu, Y., Chen, Y., Lusch, R., Chen, H., Zimbra, D., Zeng, S. 2010. User-generated content on social media: Predicting market success with online word-of-mouth. IEEE Intelligent Systems 25 (1), 75–78.

Luftman, J., Lyytinen, K., Zvi, T. Ben. 2017. Enhancing the measurement of information technology business alignment and its influence on company performance. Journal of Information Technology 32 (1), 26–46.

Maes, R. 1999. A generic framework for information management, Working paper, Department of Accountancy and Information Management, Amsterdam, Netherlands.

Malthouse, E. C., Haenlein, M., Skiera, B., Wege, E., Zhang, M. 2013. Managing customer relationships in the social media era: Introducing the Social CRM House. Journal of Interactive Marketing 27 (4), 270–280.

Malviya, S., Nair, P. S. 2016. A new approach of semi-supervised clustering with abbreviation detection and domain prediction using online dictionaries. International Journal of Engineering Science and Computing 6 (7), 1562–1567.

Miniwatts Marketing Group. 2017. Internet world stats. Available at: www.internetworldstats.com/top20.htm.

Morgan, R. E., Berthon, P. 2008. Market orientation, generative learning, innovation strategy and business performance inter-relationships in bioscience firms. Journal of Management Studies 45 (8), 1329–1353.

Mostafa, M. M. 2013. More than words: Social networks’ text mining for consumer brand sentiments. Expert Systems with Applications 40 (10), 4241–4251.

Mudambi, S. M., Schuff, D. 2010. What makes a helpful online review? A study of customer reviews on Amazon.com. MIS Quarterly 34 (1), 185–200.

Pak, A., Paroubek, P. 2010. Twitter as a corpus for sentiment analysis and opinion mining. In Proceedings of the Seventh Conference on International Language Resources and Evaluation, Valletta, Malta, pp. 1320–1326.

Palmer, J. W., Markus, M. L. 2000. The performance impacts of quick response and strategic alignment in specialty retailing. Information Systems Research 11 (3), 241–259.

Pang, B., Lee, L. 2008. Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval 2 (1–2), 1–135.

Pang, B., Lee, L., Vaithyanathan, S. 2002. Thumbs up? Sentiment classification using machine learning techniques. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, Stroudsburg, PA: Association for Computing Linguistics, pp. 79–86.

Paniagua, J., Sapena, J. 2014. Business performance and social media: Love or hate? Business Horizons 57 (6), 719–728.

Pedersen, E. R., Neergaard, P. 2009. What matters to managers? The whats, whys, and hows of corporate social responsibility in a multinational corporation. Management Decision 47 (8), 1261–1280.

Pentina, I., Bailey, A. A., Zhang, L. 2018. Exploring effects of source similarity, message valence, and receiver regulatory focus on Yelp review persuasiveness and purchase intentions. Journal of Marketing Communications 24 (2), 125–145.

Presi, C., Saridakis, C., Hartmans, S. 2014. User-generated content behaviour of the dissatisfied service customer. European Journal of Marketing 48 (9/10), 1600–1625.

Preston, R. 2014. CIO worries: Security, talent and (sadly) ‘alignment,’ Information Week, September 10.

Qiu, L., Pang, J., Lim, K. H. 2012. Effects of conflicting aggregated rating on eWOM review credibility and diagnosticity: The moderating role of review valence. Decision Support Systems 54 (1), 631–643.

Rim, H., Song, D. 2016. ‘How negative becomes less negative: Understanding the effects of comment valence and response sidedness in social media. Journal of Communication 66 (3), 475–495.

Robertson, P. L., Casali, G. L., Jacobson, D. 2012. Managing open incremental process innovation: Absorptive capacity and distributed learning. Research Policy 41 (5), 822–832.

Sabherwal, R., Kirs, P. 1994. The alignment between organizational critical success factors and information technology capability in academic institutions. Decision Sciences 25 (2), 301–330.

Sidhu, J. S., Commandeur, H. R., Volberda, H. W. 2007. The multifaceted nature of exploration and exploitation: value of supply, demand, and spatial search for innovation. Organization Science 18 (1), 20–38.
Sougiannis, T. 1994. The accounting-based valuation of corporate R&D. The Accounting Review 69 (1), 44–68.
Sparks, B. A., Browning, V. 2011. The impact of online reviews on hotel booking intentions and perception of trust. Tourism Management 32 (6), 1310–1323.
Staw, B. M., Sandelands, L. E., Dutton, J. E. 1981. Threat rigidity effects in organizational behavior: A multilevel analysis. Administrative Science Quarterly 26 (4), 501-524.
Steyvers, M., Smyth, P., Rosen-Zvi, M., Griffiths, T. 2004. Probabilistic author-topic models for information discovery. In Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York: ACM Press, pp. 306–315.
Stieglitz, S., Dang-Xuan, L. 2013. Emotions and information diffusion in social media: Sentiment of microblogs and sharing behavior. Journal of Management Information Systems 29 (4), 217–248.
Su, Q., Xu, X., Guo, H., Guo, Z., Wu, X., Zhang, X., Swen, B., Su, Z. 2008. Hidden sentiment association in Chinese web opinion mining. In Proceedings of the 17th International Conference on World Wide Web, New York: ACM Press, pp. 959–968.
Thelwall, M., Buckley, K., Paltoglou, G., Cai, D. 2010. Sentiment strength detection in short informal text. The American Society for Informational Science and Technology 61 (12), 2544–2558.
Walther, J. B., Liang, Y., Ganster, T., Wohn, D. Y., Emington, J. 2012. Online reviews, helpfulness ratings, and consumer attitudes: An extension of congruity theory to multiple sources in Web 2.0. Journal of Computer-Mediated Communication 18 (1), 97–112.
Weibo, C. 2017. Weibo reports second quarter 2017 financial results. Available at: ir.weibo.com/phoenix.zhtml?c=253076&p=irol-newsArticle&ID=2292849.
Wilson, T., Wiebe, J., Hoffmann, P. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing, Stroudsburg, Association of Computational Linguistics, pp. 347–354.
Witten, I. H., Frank, E. 2005. Data Mining: Practical Machine Learning Tools and Techniques. Oxford, UK: Elsevier Inc.
Wooldridge, J. M. 2010. Econometric Analysis of Cross Section and Panel Data. Cambridge, MA: MIT Press.
Wu, S. P. J., Straub, D. W., Liang, T. P. 2015. How information technology governance mechanisms and strategic alignment influence organizational performance: Insights from a matched survey of business and it managers. MIS Quarterly 39 (2), 497–518.
Yan, Q., Wu, S., Wang, L., Wu, P., Chen, H., Wei, G. 2016. E-WOM from e-commerce websites and social media: which will consumers adopt? Electronic Commerce Research and Applications 17, 62–73.
Yin, D., Bond, S. D., Zhang, H. 2014a. Anxious or Angry? Effects of discrete emotions on the perceived helpfulness of online reviews. MIS Quarterly 38 (2), 539–560.
Yin, G., Zhang, Q., Li, Y. 2014b. Effects of emotional valence and arousal on consumer perceptions of online review helpfulness. In Proceedings of 2014 Americas Conference on Information Systems, ATLANTA, GA: Association for Information Systems, 1–11.
Zhang, C., Zeng, D., Li, J., Wang, F. Y., Zuo, W. 2009. Sentiment analysis of Chinese documents: From sentence to document level. Journal of the American Society for Information Science and Technology 60 (12), 2474–2487.
Zheng, L., Zheng, T. 2014. Innovation through social media in the public sector: information and interactions. Government Information Quarterly 31 (Supplement 1), S106–S117.
Zhou, X., Xu, Y., Li, Y., Josang, A., Cox, C. 2012. The state-of-the-art in personalized recommender systems for social networking. Artificial Intelligence Review 37 (2), 119–132.

APPENDIX: CHINESE VS. ENGLISH TEXT MINING

Text mining is widely used in a variety of applications. Due to the characteristics of natural language, quite different text-mining techniques are needed for different languages. The following is a brief discussion on the difference between Chinese and English text mining.

As with human comprehension processes, text-mining techniques generally start with terms before building to the analysis of higher-level structures (e.g., a sentence, graph, or document). A simple tokenization process is usually effective enough in English, which has white space between sequential words. One difficulty in this case is determining the root of an inflected or derived word. Rule-based algorithms have proven effective in such processes (e.g., Lovins 1968; Porter 1980). Fortunately, Chinese words have no variant forms, so the stemming process that is so difficult in English is unnecessary in Chinese. However, because there is no white space for word segmentation in Chinese, the initial term extraction itself is challenging. The basic element of Chinese is a word. However, the basic semantic element of Chinese is a term, which is a word or a combination of up to several words. Sometimes a combination of words reads quite differently from its component words. For instance, the word 发 means "getting rich," which always has a positive orientation in sentiment analysis, and the word 福 means "luck," which is also positive. However, the combination 发福 means "becoming fat," which is generally negative. As shown in this example, improper segmentation can lead to strange and unpredictable results in Chinese text mining. Hence, segmentation is one of the most important steps in Chinese text mining. Many methods of segmentation have been applied to Chinese, including n-grams (Nie et al. 2000), conditional random fields (Peng et al. 2004) and deep learning (Zheng et al. 2013), but the problem has not been solved, especially for new-word recognition in social media.

Other problems in Chinese text mining include polysemy and synonymy. It is well known that the meaning of Chinese characters strongly relies on their context (Zhang and Jiang 2010). Frequently used terms commonly have different pronunciations and meanings in different situations; the exact pronunciation and meaning both rely on the context. Sometimes, when context information is insufficient, identification is even difficult for people, much less for algorithms. For instance, 打折 can be pronounced as "da zhe," meaning "cut price" in sales, or as "da she," meaning "broke something (e.g., a stick)" in a situation such as a fight. Due to the necessity of rhetoric, it is also common for Chinese people, even ordinary ones, to inadvertently adopt alternative expressions with the same meaning in speaking and writing. This characteristic brings serious challenges for traditional, statistics-based (e.g., lexicon-based) text mining.

Additionally, abbreviations and slang are very challenging in Chinese text mining. As in English, abbreviations are widely used in Chinese. However, because there is no uppercase in Chinese, abbreviations are usually composed of words extracted from the full name. It is quite common for such combinations to also be occupied by another term, leading to a polysemy problem. Due to Chinese's long history, slang is also widely used. The comprehension of many slang terms usually requires going beyond the context to consider, for instance, the history of the term. Existing text-mining techniques usually suffer low performance under these conditions.

Table 1. Definitions of the variables

Variable	Description
Innovation investments (RDI_{it})	The innovation investments of a specific firm *i* in a financial year *t*, which is measured by the R&D expenses of each firm in a financial year.
Revenue (RVN_{it})	The revenue of a specific firm *i* in a financial year *t*.
Valence ($SENT_{it}$)	The valence of social media posts, which is measured with the domain sentiment score of UGC for a specific firm *i* in a financial year *t*.
Volume (VOL_{it})	The volume of the social media posts for a specific firm *i* in a financial year *t*.
Researcher number ($RCHR_{it}$)	The number of researchers at a specific firm *i* in a financial year *t*.
Profit (PTF_{it})	The profit of a specific firm *i* in a financial year *t*.
Asset (AST_{it})	The assets of a specific firm *i* in a financial year *t*.
Return on assets (ROA_{it})	The return on assets of a specific firm *i* in a financial year *t*.
Employee number (EMP_{it})	The number of employees at a specific firm *i* in a financial year *t*.
Industry (IND_i)	The industry of a specific firm.
Year Dummies ($YEAR_t$)	Year dummy variables.

Table 2. Descriptive statistics for the variables

Variable	Obs	Mean	Std. Dev.	Min	Max
RDI_{it} *(log transformed)*	4430	17.610	1.390	7.984	26.144
RVN_{it} *(log transformed)*	3488	7.365	1.280	4.319	13.353
$SENT_{it}$	4428	0.067	0.028	-0.018	0.200
VOL_{it} *(log transformed)*	4430	6.247	0.614	3.689	8.051
$RCHR_{it}$ *(log transformed)*	4430	5.827	1.160	1.946	11.506
PTF_{it} *(log transformed)*	4430	18.784	1.415	12.329	24.498
AST_{it} *(log transformed)*	4430	21.861	1.130	18.811	27.318
ROA_{it}	4430	0.064	0.049	0.000	0.493
EMP_{it} *(log transformed)*	4430	7.684	1.116	3.951	12.594
IND_i	4430	Category variable. There are 96 industries			
$YEAR_t$	4430	Category variable. There are 5 years.			

Table 3. Correlations between variables

Variable	RDI_{it}	$SENT_{it}$	VOL_{it}	$RCHR_{it}$	PFT_{it}	AST_{it}	ROA_{it}	EMP_{it}	RVN_{it}
RDI_{it}	1								
$SENT_{it}$	-0.696	1							
VOL_{it}	0.299	0.004	1						
$RCHR_{it}$	0.632	-0.622	0.267	1					
PFT_{it}	0.483	-0.274	0.675	0.512	1				
AST_{it}	0.539	-0.685	0.398	0.650	0.773	1			
ROA_{it}	0.101	0.323	0.667	0.010	0.529	-0.002	1		
EMP_{it}	0.552	-0.623	0.328	0.775	0.618	0.790	0.025	1	
RVN_{it}	0.298	0.037	0.617	0.276	0.833	0.406	0.772	0.329	1

Table 4. Estimation results for the impact of valence

Variable	Random-effect Models		Fixed-effect Models	
	Model 1	Model 2	Model 3	Model 4
$SENT_{it-1}^2$	3.483*** (0.921)	4.158*** (0.890)	5.461*** (1.007)	4.823*** (1.034)
$SENT_{it-1}$	0.470 (0.596)	-0.189 (0.597)	-2.288*** (0.705)	-1.696** (0.736)
$VOL_{it-1} \times SENT_{it-1}^2$	-0.747*** (0.149)	-0.767*** (0.143)	-1.084*** (0.168)	-0.913*** (0.174)
VOL_{it-1}	0.978*** (0.040)	0.232*** (0.078)	0.455*** (0.054)	0.203** (0.100)
$VOL_{it-1} \times SENT_{it-1}$	-0.414*** (0.097)	-0.390*** (0.098)	0.303** (0.120)	0.229* (0.123)
$RCHR_{it-1}$	--	0.162*** (0.026)	--	-0.018 (0.035)
PFT_{it-1}	--	0.395*** (0.041)	--	0.007 (0.048)
AST_{it-1}	--	-0.325*** (0.049)	--	0.216** (0.075)
ROA_{it-1}	--	3.581*** (0.610)	--	1.726** (0.726)
EMP_{it-1}	--	0.023 (0.033)	--	0.214*** (0.055)
IND_i	Included but values not presented			
$YEAR_t$	--		Included but values not presented	
cons	12.848 (0.247)	15.075*** (0.720)	15.255*** (0.301)	10.214*** (1.287)
R^2: within	0.157	0.145	0.218	0.230
R^2: Between	0.666	0.823	0.503	0.454
R^2: Overall	0.592	0.727	0.430	0.421

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. According to the Hausman test, $\chi^2(10) = 609.33$, $p < 0.001$. Hence, the fixed-effect models fit our data set better than random-effect models.

Table 5. Estimation results for the impact of innovation investment

Variable	Coefficient	
	Random-effects Model	Fixed-effects Model
	Model 5	Mode 6
RDI_{it-1}	0.049***	0.004***
	(0.009)	(0.010)
$RCHR_{it-1}$	-0.018	-0.015
	(0.015)	(0.015)
PFT_{it-1}	-0.022***	0.004
	(0.008)	(0.008)
AST_{it-1}	0.707***	0.275***
	(0.021)	(0.030)
ROA_{it-1}	3.251***	3.227***
	(0.192)	(0.188)
EMP_{it-1}	0.234***	0.083***
	(0.0217)	(0.024)
cons	-10.240***	0.504
	(0.406)	(0.570)
IND_i	Included but values not presented	
$YEAR_t$	--	Included but values not presented
R^2: within	0.392	0.486
R^2: Between	0.881	0.777
R^2: Overall	0.868	0.736

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. According to the Hausman test, $\chi^2(6) = 608.07$, $p < 0.001$. Hence, the fixed-effect models fit our data set better than random-effect models.

Table 6. Estimation results for top one fifth largest firms				
Variable	Random-effects Models		Fixed-effects Models	
	Model 7	Model 8	Model 9	Model 10
$SENT_{it-1}^2$	8.373***	10.837***	8.046***	11.325***
	(2.937)	(3.238)	(2.906)	(3.161)
$SENT_{it-1}$	-10.937***	-8.977***	-9.401***	-9.784***
	(3.122)	(3.280)	(3.141)	(3.204)
$VOL_{it-1} \times SENT_{it-1}^2$	-2.434***	-2.797***	-2.206***	-2.864***
	(0.529)	(0.587)	(.527)	(0.573)
VOL_{it-1}	-0.128	0.748***	0.027	0.668**
	(0.188)	(0.271)	(0.201)	(0.265)
$VOL_{it-1} \times SENT_{it-1}$	2.524***	1.981***	2.128***	2.146***
	(0.607)	(0.641)	(0.618)	(0.627)
$RCHR_{it-1}$	--	0.139***	--	0.143***
		(0.047)		(0.046)
PFT_{it-1}	--	-0.259**	--	-0.252*
		(0.124)		(0.121)
AST_{it-1}	--	-0.073	--	-0.147
		(0.129)		(0.128)
ROA_{it-1}	--	1.571	--	1.492
		(1.429)		(1.405)
EMP_{it-1}	--	0.017	--	-0.049
		(0.818)		(0.074)
IND_i	Included but values not presented			
$YEAR_t$	--		Included but values not presented	
cons	17.735***	18.593***	16.747***	20.564***
	(0.918)	(2.082)	(1.060)	(2.091)
R^2: within	0.280	0.280	0.333	0.326
R^2: Between	0.469	0.761	0.708	0.777
R^2: Overall	0.420	0.635	0.608	0.654

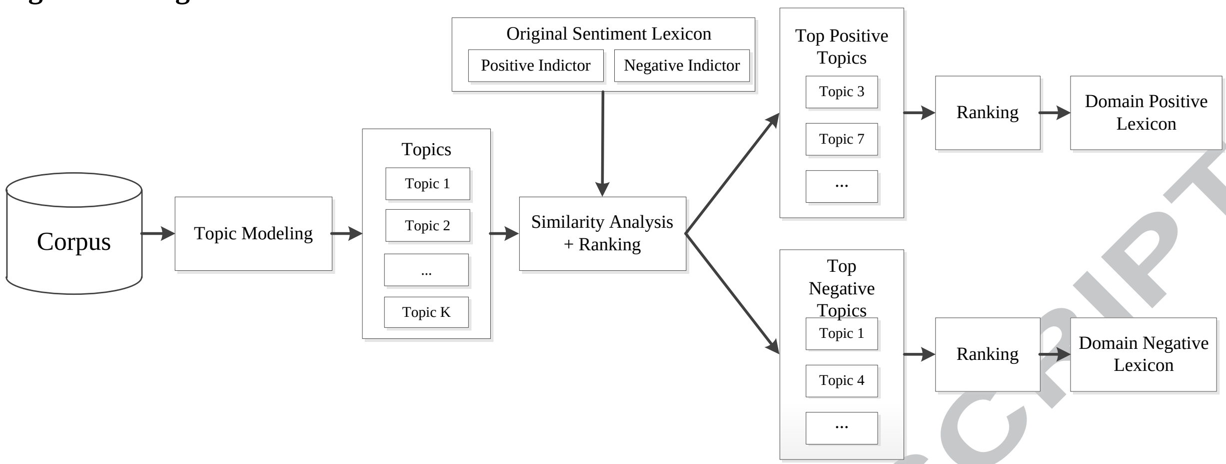
Note: * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$. According to the Hausman test, $\chi^2(10) = 105.92$, $p < 0.001$. Hence, the fixed-effect models fit our data set better than random-effect models.

Table 7. Estimation results for the 20% smallest firms

Variable	Random-effects Models		Fixed-effects Models	
	Model 11	Model 12	Model 13	Model 14
$SENT_{it-1}^2$	0.830**	5.288**	0.225**	0.839**
	(2.316)	(2.325)	(2.2505)	(2.347)
$SENT_{it-1}$	3.922***	3.597***	1.0135	1.445*
	(1.018)	(1.196)	(1.0725)	(1.291)
$VOL_{it-1} \times SENT_{it-1}^2$	-0.169**	-0.825***	-0.121**	-0.175**
	(0.354)	(0.354)	(0.351)	(0.364)
VOL_{it-1}	0.894***	0.182	0.413***	0.171
	(0.088)	(0.124)	(0.093)	(0.160)
$VOL_{it-1} \times SENT_{it-1}$	-0.925***	-0.910***	-0.222	-0.300*
	(0.157)	(0.186)	(0.173)	(0.205)
$RCHR_{it-1}$	--	0.173***	--	0.034*
		(0.053)		(0.064)
PFT_{it-1}	--	0.334***	--	0.093*
		(0.064)		(0.068)
AST_{it-1}	--	-0.031	--	0.139
		(0.092)		(0.146)
ROA_{it-1}	--	3.562***	--	0.960*
		(1.124)		(1.361)
EMP_{it-1}	--	-0.204***	--	-0.069
		(0.076)		(0.101)
IND_i	Included but values not presented			
$YEAR_t$	--		Included but values not presented	
Cons	13.320***	12.798***	16.088***	12.924***
	(0.581)	(1.357)	(0.596)	(2.549)
R^2: within	0.202	0.204	0.293	0.300
R^2: Between	0.709	0.884	0.606	0.556
R^2: Overall	0.645	0.820	0.383	0.483

Note: * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$. According to the Hausman test, $\chi^2(10) = 188.07$, $p < 0.001$. Hence, the fixed-effect models fit our data set better than random-effect models.

Figure 1. Design and Process



Highlights

Theorizes about the impact of valence and volume of UGC on firm innovation investment.
Empirical study conducted with 886 listed firms and their relevant 6.2 million microblogs.
Advancing sentiment analysis done by constructing timely domain-specific sentiment lexicons.
Valence of user-generated content has a U-shaped relation with innovation investment.
Volume of user-generated content reduced the impact of valence.