# MORE THAN THE TONE: THE IMPACT OF SOCIAL MEDIA OPINIONS ON INNOVATION INVESTMENTS

*Research in Progress*

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

Lele, Kang, Nanjing University, Nanjing, China, lelekang@nju.edu.cn

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

Yiyang, Bian, City University of Hong Kong, Hong Kong SAR, China, bianyi-yang321@gmail.com

## Abstract

*Social media is a valuable knowledge source for firm innovation. Extending the literature of both social media and innovation management, we attempt to examine how the valence and volume of user-generated content (UGC) from social media influence firm organizational innovation behaviours. In this research-in-progress study, we have reviewed the existing literatures and proposed three hypotheses. Firstly, we propose that valence of UGC from social media has a U-shaped relation with firm innovation investments. In particular, compared with neutral UGC, both negative and positive contents are found to push firms to invest more in innovation. Secondly, we argued that such a curvilinear relation is mitigated with an increase in volume of UGC. Last but not least, we argued that firm investment in innovation improves firm performance. To validate our proposed hypotheses, we have designed an innovative framework of sentiment analysis and collected a large dataset including 5-year panel with 886 listed firms and their relevant 6.2 million micro-blogs. The preliminary results from applying sentiment analysis into the collected dataset are reported in this study. In the future, we will validate our hypotheses with more sophisticated estimation models and strict robustness check. The potential contribution to theory and practice is also discussed.*

*Keywords: Social media, Innovation investment, Valence, Volume, Firm performance*

# 1 Introduction

The business models or so-called revenue models of most social media companies, such as Facebook, Twitter, or Weibo, are similar and intuitive. Business customers are willing to pay because social media helps them build up a communication channel with their customers. In particular, those firms—the business customers—can rely on social media not only to intensify the conventional dialog between their customers but also create new ways of gaining insights from customers drawn from customer-to-customer interactions (Libai et al., 2010).

By investigating user-generated content (UGC) in social media, firms can gain insights and opinions directly from the market (i.e., from their own customers, competitors, and even their competitors' customers) to improve their business performance (He et al., 2013). Collectively, the positive role of social media in reshaping businesses is not doubted. However, transforming a UGC data set into a business asset is highly contingent upon the research context, which has resulted in non-unified metrics and implementation in prior studies. The effective use of social media for business has been discussed from three perspectives in previous studies. The first stream can be categorized as research with a design science perspective, which attempts to develop and implement methods or tools specializing in analyzing UGC from social media (Aral et al., 2013). Although such descriptive output allows users to interpret public opinions in an intuitive way, the underlying mechanism between UGC and business performance or prospective implications is not well understood due to a lack of understanding of causality. The second stream concentrates on the business logic characterized by social media, such as the role of social media in reshaping conventional business (Paniagua and Sapena, 2014; Yan et al., 2016) or the business value of the customer opinions or engagement (Mudambi and Schuff, 2010).The findings from above studies provide insightful theoretical implications, the measurement used in these studies, such as counting the number of words of UGC, are unable to remove the noise or biases existing in large-scale UGC. In other words, the methods used in previously studies have not been sophisticated enough to completely resolve the business problem. The last stream in this domain can be regarded as an intermediate between the aforementioned two approaches in which fundamental analytical methods have been applied to understand Internet-driven or stylish 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 penetrate into the niche domain in which business can be seamlessly integrated with social media. In other words, such findings cannot provide unified solutions to firms across different industries. To this end, it is imperative to propose a relatively generalizable approach to understand how to squeeze business values from social media.

In this study, we try to employ multisource data sets comprising UGC from the largest micro-blogging service provider in China (Weibo) as well as the financial and operational information from 886 listed firms in China. We attempt to understand how public opinions in social media influence the degree of innovation investments and the consequences of such investments. Our work may contribute to the existing literatures in three ways. First, we try to propose an advanced sentiment analysis method that reveals a curvilinear (i.e., a U-shaped) relation between valence in the public opinions and innovation investments. Second, we try to figure out whether the volume of public opinions will moderate the aforementioned curvilinear relation or not. We also want to understand whether the volume of public opinions mitigate the influence of valence on innovation investments. Third, our research may echo the positive association between investment in innovation and firms performances in terms of financial indicators (Belderbos et al., 2004; Grabowski and Vernon, 1990).

# 2 Literature Review

Many studies have focused on the value of customer reviews (i.e., typical UGC for a product) by simply counting the number of words in a review or quantifying the review ratings (Chevalier and Mayzlin, 2006; Mudambi and Schuff, 2010). These methods fail to extract core values (i.e., customer opinions) within the massive data set. Therefore, cutting-edge tools to transform UGC from social media into business assets are necessary, particularly for opinion-oriented knowledge from crowds (Kane et al., 2014; Yates and Paquette, 2011). In this regard, sentiment analysis, evolved from TM (text mining) or NLP (natural language processing) techniques, has been applied to analyze such free-form UGC.

In a lexicon-based approach, predefined sentiment lexicons that contain both positive indicators (term) and negative indicators are adopted to extract the number of corresponding sentiment indicators based on string matching (Pang and Lee, 2008). Due to its intuitive outputs (e.g., a large number of positive terms indicates a strong positive orientation), the lexicon-based approach has been widely adopted in business practice and research (Liu et al., 2010). However, this method may suffer from low recall because it strongly relies on the completeness of sentiment lexicons. Furthermore, it may lead to confusion when facing synonymy and polysemy issues. In an ML-based approach, a set of labeled data (training data) is used to train the classifier to learn the “rules” (Witten and Frank, 2005). Then, the trained classifiers are used to predict the unlabeled data based on the “rules” they learned (Pak and Paroubek, 2010; Pang et al., 2002). Such a process is generally carried on a text fragment level (e.g. sentence, document) instead of a word level. However, the overall learning rules or prediction processes are metaphorically described as a “black box” to users, which results in a dilemma when it comes to explaining or improving the algorithm. In a linguistic approach, researchers have attempted to understand the semantic meaning of text and draw conclusions based on this meaning (Wilson et al., 2005). Such an approach is similar to the process of human cognition. However, due to the complexity and flexibility of human language (e.g., negation, idioms), this approach is not easily implemented in real-world applications. In this study, we incorporated the widely used lexicon-based approach and adopted an automatic lexicon expansion method to improve the completeness of the lexicon. Moreover, our new method constructs and updates the lexicon based on domain knowledge, which, to some extent, alleviates the synonymy and polysemy problems.

In information systems studies, sentiment analysis is employed as a tool or application to resolve business or societal questions. Mostafa (2013) employed an expert-predefined lexicon to create an application with sentiment analysis, which was then used to investigate hidden patterns in consumers’ opinions toward international brands. Stieglitz and Dang-Xuan (2013) directly used a commercial software package, “SentiStrength”[1], to analyze the level of sentiments in politically relevant UGC from Twitter. These authors found that emotionally charged information was correlated with online propagation in terms of frequency and velocity of the online transmission. Goh et al. (2013) applied sentiment analysis to investigate both marketer-generated content (MGC) and UGC in a fan page brand community on Facebook and found that valence in UGC had a stronger influence than MGC on consumer-purchasing behaviors. These findings provide sound evidence that sentiment analysis is an effective tool for investigating how opinions expressed from UGC play a role in decision-making.

Besides the contribution to online marketing and community, UGC from social media serves as an important supplement to product or service innovation. Prior literature has highlighted how to cultivate internal innovation by relying on external information, especially knowledge contributed by large number of users (Chatterji and Fabrizio, 2014; Garriga et al., 2013; Laursen and Salter, 2006). We believe that users’ opinions regarding products or service can be transformed into more fine-grained and comprehensive knowledge from the their posts in social media with the support of TM techniques, particularly sentiment analysis. More specifically, encouraged by a study by Liu and colleagues (2010), we used volume, which denotes the amount of communication, and valences, which denote the sentiment of that content, as two important indicators to predict how the strategic alignment of innovation investments is influenced by UGC. Our hypotheses are developed in the next section.

[1] http://sentistrength.wlv.ac.uk/

# 3 Hypothesis Development

## 3.1 Valence of Opinions in UGC

The valence is used to present the overall orientation (i.e., positive or negative) and intensity (i.e., weak or strong) of sentiments expressed by texts (Lerner and Keltner, 2000). Whether positive or negative expression has more influence on individuals is still under debate in the current literature (Derks et al., 2008; Liu et al., 2010; Yin et al., 2014). The proponents favoring the influential role of negative text have argued that negative stimuli tend to elicit stronger and faster reactions than neutral or positive stimuli (Baumeister et al., 2001). Such an argument is evidently supported by findings in online settings. For instance, negative information is found to have a greater influence on people's behavior than neutral or positive information (Yin et al., 2014); in social media, negative information is found to be more liable to trigger arousal and curiosity and facilitate information diffusion (Kimmel and Kitchen, 2014). Therefore, in a similar vein, compared with information with a neutral or positive slant, the contents or points expressed in a negative manner can be more easily noted by focal firms. Negatively expressed information will be assigned a higher weight in making organizational decisions, such as the alignment of innovation investments. Furthermore, prior studies in the fields of both marketing and IS have argued that individuals tend to provide evaluative opinions with negative expressions (Derks et al., 2008; Liu et al., 2010; Yin et al., 2014). Last but not least, it has been also found that firms feel more pressure when negative UGC exists in social media (Pedersen and Neergaard, 2009). Such a finding is not counterintuitive to common sense. Due to the rapid propagation of information in social media, negative UGC regarding certain products or services may influence not only a firm's own customers but its potential customers as well (Walther et al., 2012). As a result, pressure will push a firm to improve its products or services in terms of innovation, which will consequently influence its decisions in terms of innovation investments. Determining whether positive or negative expression has a stronger influence on individual behavior is still under debate. Although it seems that UGC expressed in a negative manner can attract attention intuitively, individuals or organizations are not only influenced by negatively expressed content. User-generated content expressed positively can arouse attention as well. Anchored in attention theory, an external information search is a controlled and proactive process that involves attention, examination, and assessment (Li et al., 2013). Collectively, based on the deduction noted above, we find that both negative and positive UGC can attract firms and encourage them to make innovation investments. Therefore, we postulate our first hypothesis:

> ***Hypothesis 1***: *The valence of UGC (to a firm) in social media has a significant quadratic relation (U-shape) with its innovation investments.*

## 3.2 Moderating Role of Volume

Valence reveals the overall sentiment of UGC in social media, and volume measures the amount of UGC (Chen et al., 2011; Liu et al., 2010). The volume of UGC has received the same level of attentions as valences in prior studies (Etzion and Awad, 2007; Flanagin and Metzger, 2013). In this study, we argue that the volume of UGC plays a moderating role in the relation between valence and investment in innovation. In particular, the quadratic relation proposed in H1 will be mitigated with an increased volume of UGC. Such an argument is supported by two arguments from the perspective of strategic alignment of resources, which are outlined below. For negative UGC, a high volume may reflect the complaining or unsatisfactory voice of the market to the focal firm or its products or services (Presi et al., 2014). With an increase in the volume of negatively expressed UGC, corporations may face a series of challenges, such as customer churn (Malthouse et al., 2013), reputation risk (Aula, 2010), and discouragement of investors (He et al., 2013). In this regard, a firm at a point of crisis should allocate its resources to resolve the immediate dilemma. Continuing to increase investments in innovation may not be a wise decision. However, considerable amounts of positive UGC may also deter subsequent investment in innovation. Positively expressed UGC signals that the public is praising the current performance of the focal firm (He et al., 2013). Therefore, the firm may wish to maintain its current status or strategies, or at best make slight improvements or adjustments, rather than making huge investment to create disruptive innovations. In general, counterintuitively, we argue a high volume of UGC with either positive or negative expression does not contribute to an increase in innovation-related investments. Therefore, the following hypothesis is proposed:

> ***Hypothesis 2:*** *The quadratic relation between valence of UGC (to a firm in social media) and innovation investments is mitigated by an increasing volume of UGC.*

### 3.3 The Consequences of Innovation Investments

The existing literature repeatedly includes evidence supporting the positive role of innovation investments in improving financial performance (Belderbos et al., 2004; Grabowski and Vernon, 1990; Levin, 1988; Sougiannis, 1994). Generally, innovation is vital for a firm—regardless of its position in the market—to maintain its competitive advantages and defeat its rivals (Cooke and Wills, 1999; Morgan and Berthon, 2008). Innovation is associated with a series of activities through which a firm gains innovative knowledge about its products or services. In a fiercely competitive market, breakthrough products or services can help firms attract more customers (Katila and Ahuja, 2002) and consequently outperform their rivals. Open innovation theory also predicts that increasing reliance on external innovation activities is beneficial for firm financial performance ( Chesbrough et al., 2006). When firms increase innovation investments based on UGC accessed from the market, these inter-corporation linkages allow firms to keep up with evolving products or services. Therefore, firms can gain tangible benefits from innovation investments (Berchicci, 2013). Given that innovation investments are a precondition of productive output (DiMasi et al., 2003; George et al., 2002), it is reasonable to make the assumption that there is a positive relation between innovation investments and later business performance. Therefore, the following hypothesis is proposed:

> ***Hypothesis 3:*** *Innovation investments increase a firm's performance.*

## 4 Research Method

### 4.1 Framework for Sentiment Analysis

We conducted sentiment analysis to determine public opinion embedded in the UGC analyzed in this study. In particular, the valence of the sentiment was measured using a sentiment score that summarizes people's orientation to a specific firm (e.g., its products or services). The sentiment embedded in a text (e.g., a document or microblog) can be mined by means of sentiment analysis techniques (Pang and Lee, 2008). We adopted a topic modeling method to expand the existing sentiment lexicon to construct timely domain-specific sentiment lexicons. In the topic modeling process (i.e., Latent Dirichlet Allocation, LDA), each text segment (i.e., UGC in a microblog in our context) is assumed to be characterized by a mixture of latent topics, and each topic can be characterized by a mixture of terms. After a series of testing rounds, the perplexity score with the lowest value was selected. In particular, the perplexity score of a corpus $D$ was defined 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) that composes a set of words. The conditional probability $Pr(d|\theta,\varphi)$ is the generation probability of all the terms $w \in d$.

We employed an original sentiment lexicon as seeding. In terms of treating each topic as a bag of words, we calculated its distance with the original positive lexicon and original negative lexicon using similarity analysis. Given that there is noise in those words compositing each topic, a supplementary procedure was executed. The Chinese sentiment analysis lexicon HowNet20 (Dong and Dong, 2003) was deployed as the original sentiment lexicon for this work. The design of our sentiment analysis and the entire work-flow process is illustrated in Figure 1.

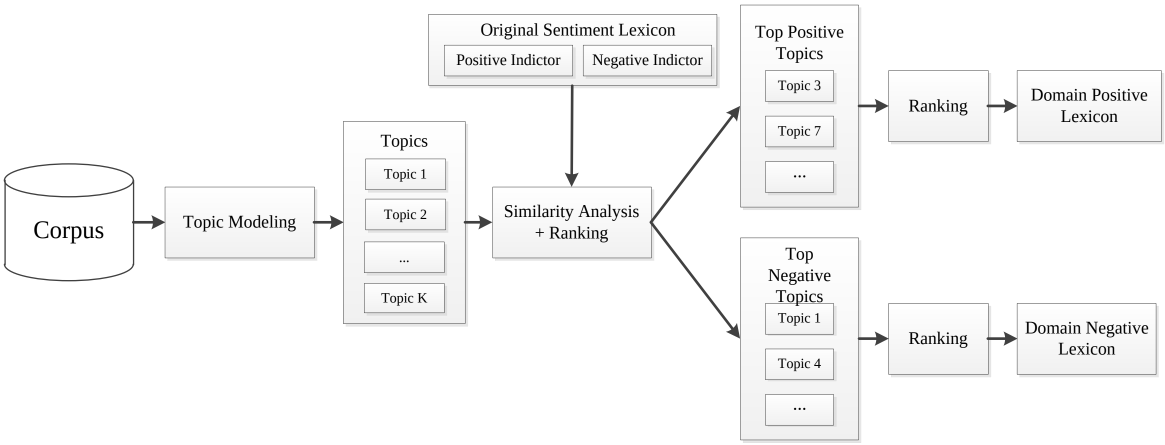


*Figure 1. Design process*

After constructing the domain sentiment lexicon, we extracted the domain positive words and domain negative words in the UGC relating to each firm. 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 domain positive words and domain negative words respectively.

## 4.2 Research Setting and Measurement

In this study, we focused on Chinese firms, particularly A share-listed firms in the Shanghai and Shenzhen stock markets Both Compustat and GTA[2] are used to depict firm demographics and performance. Compustat was used to obtain financial indicators, and GTA was used to obtain information about innovation investments and firm demographics. Finally, 886 firms, which 1) overlapped in both databases and 2) are free of missing values, were selected to construct our strictly balanced panel (i.e., 4430 observations from 2011–2015).

The UGC from social media was collected from Weibo[3], which is the largest social media platform in China. We used firm names as the keywords for data crawling through the Weibo API. Given that UGC is free-form data with many abbreviations, we adopted abbreviation-detection methods (Kiss and Strunk, 2006; Malviya et al., 2016) to identify and construct the abbreviation set for each firm.

[2] http://www.gtafe.com/

[3] https://weibo.com/

| Variable | Description |
|---|---|
| Innovation investments (*RDI*) | The innovation investments of a firm, which is measured by the R&D expenses of each firm in a financial year. |
| Revenue (*RVN*) | The revenue of a firm in a financial year. |
| Valence (*SENT*) | The valence of social media posts is measured with the domain sentiment score of UGC for a firm in a financial year. |
| Volume (*VOL*) | The volume of the social media posts for a firm in a financial year. |
| Researcher number (*RCHR*) | The number of researchers at a firm in a financial year. |
| Profit (*PTF*) | The profit of a firm in a financial year. |
| Asset (*AST*) | The assets of a firm in a financial year. |
| Return on assets (*ROA*) | Return on assets of a firm in a financial year. |
| Employee number (*EMP*) | The number of employees at a firm in a financial year. |
| Industry (*IND*) | The industry of a firm. |

*Table 1. Description of the variables.*

The definitions and descriptions of all of the variables are summarized in Tables 1 and Table 2. To mitigate the possibility of reverse causality, all independent and control variables will be lagged one year in the future estimations. Due to the variation in the firms' sizes in the sample, we transformed all firm-size-related variables using logarithms.

| Variable | Obs | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| *RDI (log transformed)* | 4430 | 17.610 | 1.390 | 7.984 | 26.144 |
| *RVN (log transformed)* | 3488 | 7.365 | 1.280 | 4.319 | 13.353 |
| *SENT* | 4428 | 0.067 | 0.028 | -0.018 | 0.200 |
| *VOL (log transformed)* | 4430 | 6.247 | 0.614 | 3.689 | 8.051 |
| *RCHR (log transformed)* | 4430 | 5.827 | 1.160 | 1.946 | 11.506 |
| *PTF (log transformed)* | 4430 | 18.784 | 1.415 | 12.329 | 24.498 |
| *AST (log transformed)* | 4430 | 21.861 | 1.130 | 18.811 | 27.318 |
| *ROA* | 4430 | 0.064 | 0.049 | 0.000 | 0.493 |
| *EMP (log transformed)* | 4430 | 7.684 | 1.116 | 3.951 | 12.594 |
| *IND* | 4430 | Category variable. There are 96 industries | | | |

*Table 2. Basic statistics of the variables*

## 4.3 Estimation Procedure

To validate our hypotheses, we will develop empirical models to assess the explanatory power of each variable. We will firstly examine the influence of UGC from social media on innovation investments to validate Hypotheses 1 and 2. Specifically, we will regress innovation investments on the squared term of valence as well as the interaction term between volume and the squared valence. Other terms, such as valence and the interaction term between valence and volume, will also be included in the models, as suggested by Wooldridge (2010). We will apply both random-effect and fixed-effect models to validate our hypothesis.

# 5 Discussions and Conclusions

This study will make three potentially important contributions to both IS and the innovation literature. First, we propose and validate a novel framework of sentiment analysis. Second, our work expands the theoretical boundary of the impact of social media on the strategic alignment of innovation investments. We postulate that the opinions expressed from the crowds in the social media can exert impacts on firms' innovation decision. The negative expression can motivate firms to increase the innovation investment. In contrast, the positive ones signify the firms to continue the current maneuver in innovation investment for further outperformance. In Third, our work contributes to the social media literature by determining a quadratic relation between valence and innovation investments. The finding will provide evidence that public opinions expressed through social media do influence firms' decisions pertaining to innovation investments.

Our findings will also afford two important implications for practitioners. First, we provide a clear roadmap to implementing a novel framework of sentiment analysis to mine public opinions. Practitioners can utilize our design to implement our proposed method to analyze public opinions regarding the focal companies. Second, our findings provide an alternative approach for firms to make decisions about innovation investments.

As a research-in-process study, we look forward to receiving valuable comments and suggestions to improve this work in the future. In particular, we think the limitations may exist in the following three aspects. Firstly, although the Chinese sentiment analysis lexicon HowNet20 has been widely accepted as one of the best candidates, we still believe that there is room for improvement, particularly for further refining the qualitative degrees of adjectives in Chinese. Secondly, although we will use various methods such as cross-validated topic modeling for managing the contextual meaning in opinion mining and increased the accuracy rate to an acceptable rate, tools or methods with more sophisticated methods for understanding linguistic contexts should be employed in future studies. Thus, more sophisticated approach ought to be employed for significantly reducing potential bias in the sample selection. Thirdly, we will review more relevant studies from the business aspect in the future. In current work, we primarily discussed the relevant literatures from analytical aspect. Last but not least, the moderating effect should be further developed in accordance with the overarching theoretical framework in the future work.

# 6 Acknowledgement

The study was supported by the following grants: the National Natural Science Foundation of China (NSFC 71704078, NSFC 71702133, NSFC 71532015, NSFC 71771177, NSFC 71771179), the Innovation Program of Shanghai Municipal Education Commission (15CG20), the Shanghai Pujiang Program (16PJC086), Beijing brain research project of Beijing Municipal Science & Technology Commission (No.Z171100000117009), and the Fundamental Research Funds for the Central Universities and the Research Funds of Renmin University of China (No. 17XNLF05).

## References

Aral, S. Dellarocas, C. and D. Godes (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.

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

Belderbos, R. Carree, M. and B. Lokshin (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., and F. Okumus (2016). “Understanding satisfied and dissatisfied hotel customers: Text mining of online hotel reviews.” *Journal of Hospitality Marketing and Management* 25 (1), 1–24.

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

Chen, Y. Wang, Q., and J. Xie (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. and J. West (2006). *Open innovation: Researching a new paradigm*. Oxford University Press.

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

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

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

Dong, Z. and Q. Dong (2003). “HowNet - A hybrid language and knowledge resource.” In: *Proceedings of* NLP-KE 2003 - 2003 International Conference on Natural Language Processing and Knowledge Engineering, p. 820–824.

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

Etzion, H. and N. F. Awad (2007). “Pump up the volume? Examining the relationship between number of online reviews and sales: Is more necessarily better? In: *Proceedings of* International Conference on Information Systems 2007, p. 1–14.

Flanagin, A. J. and M. J. Metzger (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.

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

George, G. Zahra, S. A. and D. R. Wood (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.

Goh, K. Y. Heng, C. S. and Z. Lin (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. and J. Vernon (1990). "A new look at the returns and risks to pharmaceutical R&D." *Management Science* 36 (7), 804–821.

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

Kane, G., Labianca, G. and S. P. Borgatti (2014). "What's different about social media networks? A framework and research agenda." *MIS Quarterly* 38 (1), 274–304.

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

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

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

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

Laursen, K. and A. Salter (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. and D. Keltner (2000). "Beyond valence: Toward a model of emotion-specific influences on judgement 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.

Li, Q. Maggitti, P. G. Smith, K. G. and R. Katila (2013). "Top management attention to innovation: The role of search selection and intensity in new product introductions." *Academy of Management Journal* 56 (3), 893–916.

Libai, B. Bolton, R. Bügel, M. S. de Ruyter, K. Götz, O. Risselada, H. and A. T. Stephen (2010). "Customer-to-customer interactions: Broadening the scope of word of mouth research." *Journal of Service Research* 13 (3), 267–282.

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

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

Malviya, S. and P. S. Nair (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.

Morgan, R. E. and P. Berthon (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. and D. Schuff (2010). "What makes a helpful online review? A study of customer reviews on Amazon.com." *MIS Quarterly* 34 (1), 185–200.

Pak, A. and P. Paroubek (2010). "Twitter as a corpus for sentiment analysis and opinion mining. In: *Proceedings of* the Seventh Conference on International Language Resources and Evaluation, p. 1320–1326.

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

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

Pang, B. Lee, L. and S. Vaithyanathan (2002). “Thumbs up?: Sentiment classification using machine learning techniques.” In: *Proceedings of the* Conference on Empirical Methods in Natural Language Processing, p. 79–86.

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

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

Sougiannis, T. (1994). “The accounting based valuation of corporate R&D.” *The Accounting Review* 69 (1), 44–68.

Stieglitz, S. and L. Dang-Xuan (2013). “Emotions and information diffusion in social media—Sentiment of microblogs and sharing behavior.” *Journal of Management Information Systems* 29 (4), 217–248.

Walther, J. B. Liang, Y. Ganster, T. Wohn, D. Y. and J. Emington (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.

Witten, I. H. and E. Frank (2005). *Data Mining: Practical Machine Learning Tools and Techniques.* Oxford, UK: Elsevier Inc.

Wilson, T. Wiebe, J., and P. Hoffmann (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*, p. 347–354.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT Press.

Yan, Q. Wu, S. Wang, L. Wu, P. Chen, H. and G. Wei (2016). “E-WOM from e-commerce websites and social media: Which will consumers adopt?” *Electronic Commerce Research and Applications* 17, 62–73.

Yates, D. and S. Paquette (2011). “Emergency knowledge management and social media technologies: A case study of the 2010 Haitian earthquake.” *International Journal of Information Management* 31 (1), 6–13.

Yin, D. Bond, S. D. and H. Zhang (2014). “Anxious or angry? Effects of discrete emotions on the perceived helpfulness of online reviews.” *MIS Quarterly* 38 (2), 539–560.