 This is the Preprint Version.

How Does Content and Sentiment Coherence Influence Online Discussion?

Short Paper

Alexis Chen
The Hong Kong Polytechnic University
Hung Hom, Kowloon, Hong Kong
zihan.chen@connect.polyu.hk

Yue (Katherine) Feng
The Hong Kong Polytechnic University
Hung Hom, Kowloon, Hong Kong
katherine.feng@polyu.edu.hk

Abstract

While online discussion has been extensively studied in the previous literature, the role of information coherence along the course of discussion is less being explored. In this study, we investigate the influence of information coherence in online discussion platforms with a focus on the sequential patterns of user posts. We measure the textual information by two dimensions: content and sentiment. Using the data from two popular online automobile discussion platforms, we empirically estimate the dependency of content and sentiment coherence among user posts and evaluate the effect of information coherence on user participation in the online discussion. The empirical evidence sheds light on the heterogeneity of the coherence influence and reveal the synergetic relationships among information coherence, dependency, and topics in the online discussion platforms. Our findings provide insights into online discussion platforms and business managers on their strategic designs to keep users engaged and facilitate digital collaboration.

Keywords: information coherence, dependency, online discussion, sequence, text mining

Introduction

Online discussion communities have become prominently prevalent in our daily life and facilitated information exchange among online users and other related parties. According to a most recent report[1], more than half of the world population have participated in social media discussions as of July 2020. Both firms and customers can benefit from the huge volume of user interactions via online discussion (Manchanda et al. 2015; Wang and Chaudhry 2018).

While many studies have investigated the generation of online discussion from multiple perspectives, ranging from social influence (e.g., Dewan et al. 2017; Wang et al. 2018), user characteristics (e.g., Kane et al. 2014; Wasko and Faraj 2005), and firm interventions (e.g., Ma et al. 2015; Wang and Chaudhry 2018), the role of information coherence along the course of discussion has been seriously underestimated. There is little evidence documenting the information coherence patterns and their impacts on online discussion. Different from the contexts of social media where users join the activities with a key motivation to build social connections, online discussion platforms are more dedicated to the function of information exchange with a unique feature of displaying user posts with the sequence. In other words, the sequential patterns of user posts are highlighted such that the latter users will read the former information before they reply. In this study, we aim to bridge the gap to understand the dynamics of online discussion with a focus on the role of information coherence along the sequence of user posts.

[1] https://datareportal.com/reports/digital-2020-july-global-statshot

There is no consensus in the literature on information coherence. Some studies found that people prefer opinion-consistent information (Ehrlich et al. 1957), while others suggested that people tend to participate in controversial discussions (Chen and Berger 2013). Some existing studies explained the findings due to individual preference or memory ability by laboratory experiments (Kleck and Wheaton 1967; Wyer Jr 1970). However, little research has been devoted to investigating the impact of information coherence in the context of online discussion platforms based on the linguistic features of user posts. In this paper, we aim to understand the influence of information coherence in terms of content and sentiment on online discussion. Particularly, we emphasized the sequential patterns among the user posts and estimated the dependency therein. The influence of information coherence may further depend on the degree of dependency as well as different topics.

Our empirical analyses were conducted using the data from two of the leading online automobile discussion platforms in China. We used text mining methods to derive the information coherence measures and to infer the topics. Stage 1 analysis investigates the coherence patterns and estimates the dependency therein. In stage 2, we analyzed the effect of information coherence on online discussion, contingent on the dependency and topics. Our preliminary results show that: 1) users are motivated to participate in the discussion when content is coherent, while their participation will decrease with higher coherence in sentiment; 2) the interactions of information coherence with dependency and topics suggest a nuanced picture that a higher degree of dependency may enhance users' perception of redundancy under coherent information and thus hinder the discussion, and there is a similar negative interacting effect with content coherence in the discussing threads of the topic concerning information acquisition.

Our study contributes to the existing literature in several aspects. First, we consider information coherence as a key antecedent of online discussion, given the unique feature of sequence in online discussion platforms. Second, we uncover the coherence patterns and the dependency, based on the textual information measured by two dimensions: content coherence and sentiment coherence. Third, we evaluate the effect of information coherence on online discussion and reveal the synergetic relationships among information coherence, dependency, and topics. Our findings provide insights into online discussion platforms and business managers on their strategic designs to keep users engaged and facilitate digital collaboration.

Literature Review

Dynamics of Online Discussion

Online discussion denotes that users exchange opinions, feelings, or experience on a topic in online communities. A stream of literature has focused on the generation of online discussion (or UGC: user-generated content) through different perspectives and emphasized the dynamics of online discussion such that subsequent user posts are influenced by previous information (Moe and Trusov 2011; Wang et al. 2018). From a social influence perspective, Wang et al. (2018) investigated how friends tie influence online ratings. They found that online users are socially nudged and tend to follow friends' previous patterns. Recent papers have investigated the effect of firm interventions on subsequent opinions. For example, Wang and Chaudhry (2018) studied how managers' responses to previous users' reviews affect subsequent reviews. They calculated the sentiment of users' reviews and the similarity of managers' responses to the reviews. Their findings suggest that managers' responses to the negative reviews help to fix problems, thus leading to a positive impact on subsequent opinions. However, those responses to the positive reviews are treated as deliberate promotion and result in a lower subsequent opinion. While information similarity in the review threads has been underlined, their analysis is from the firms' standpoint and applied to the platforms which encourage firms to respond to customers' feedback.

Moreover, in the online review contexts, plenty of studies have only used the aggregated measures from rating information, such as rating volume, valence, and variance (Dellarocas et al. 2007), and investigated their impacts on either business outcomes or rating generation (Chevalier and Mayzlin 2006). The sequential pattern along the course of user interactions is largely underexplored. Only a few exceptions have considered the sequential factor. Godes and Silva (2012) explicitly pointed out that previous research has neglected the inter-correlations across the online posts. However, their empirical examination was conducted essentially the same as other studies, i.e., using a total number of previous ratings to indicate a sequence effect. Besides, users on pure review platforms are less likely to interact with each other, as their reviews can be written based on their own product experience in many cases. In contrast, reading previous posts by sequence is an inevitable endeavor and thus becomes the unique feature of participating in the online discussion. Kane et al. (2014) investigated topic evolvement in the online discussion forum and found that users' roles and production focus vary across the stages of topic development. Their analysis, however, did not demonstrate how information change over time influence user participation in the online discussion.

In our work, we consider coherence based on the sequence of user posts. We aim to investigate the effect of information coherence on online discussion in the context where users are self-motivated to participate, mainly for the purposes to exchange their knowledge of a product of interest or to share personal experience of using the product. Moreover, the users in our context are mostly acquaintances with common interests in a product but with less social connections. Therefore, social influence plays a limited role in our study.

Information Coherence

Coherence has been defined as "a continuity of sense" (De Beaugrande and Dressler 1981). It plays a prominent role in both (text-)linguistic and psycholinguistic theories regarding text and discourse (Geeraerts and Cuyckens 2007). While some researchers found that people prefer opinion-consistent information (Ehrlich et al. 1957), others supported that people are willing to be exposed to inconsistent information (Feather 1962) and participate in controversial discussions (Chen and Berger 2013). Based on the laboratory experiments, previous studies explained the contingent findings due to individual preference and memory ability (Kleck and Wheaton 1967; Wyer Jr 1970). However, whether and to what extent the coherence influences discussion in online platforms remains unclear. In this study, we investigate the information coherence in the online discussion based on the linguistic features of user posts. Information coherence is defined as the information continuity conveyed over the sequence of user posts.

Following the literature, we incorporate two dimensions of information coherence. The first dimension is the content coherence. Streufert (1973) conducted laboratory experiments and showed that receiving the relevant information to some events can positively influence the volume of group decisions making under complex environments. Cheung et al. (2008) found that the content of electronic word-of-mouth is a key factor in adopting online opinions. In this study, we use content coherence to indicate the extent to which the content among user posts in the discussion is relevant to each other. The second dimension is the sentiment coherence. Villarroel Ordenes et al. (2017) investigated the sentiment incoherence across sentences within text-based reviews. They found that an increase of sentiment incoherence has a negative effect on the overall sentiment strength, because the sentiment expresses a set of sequentially organized propositions to explain an overall opinion and the use of contradictory sentiment expressions might convey a less degree of conviction. A similar point has been raised by Das and Chen (2007) that studies cannot disregard patterns of sentiment across sentences. In our study, we use sentiment coherence to indicate the extent to which the sentiment among user posts in the discussion is consistent with each other.

We investigate the information coherence by text mining techniques and measure the information coherence in terms of content and sentiment, as the proxies of content relevance and sentiment consistency in the online discussion. Leveraging a forecasting algorithm, we fit the patterns of coherence along the sequence of user posts and uncover the dependency therein. We further analyze the effects of content and sentiment coherence on user participation in the online discussion.

Empirical Analysis

Research Setting and Data

We collected the data from two of the leading online platforms which are dedicated to automobile discussion in China. The two platforms provide similar discussion functions to the users who are interested in automobiles. Both platforms contain more than 3 million daily active users on average, with one platform having a relatively larger user base than the other. Like all the online discussion platforms, a user initiates a topic by an initial post and other users (including the initiator) post and reply under this topic. The collection of posts is called a thread and displayed by time sequence.

Our dataset consists of 31,507 threads related to one automobile brand from both platforms between January 2018 and October 2018. Among them, 2,421 threads have fewer than five posts, which are not long enough to detect sequential patterns and were removed from the final sample. We also removed the threads that contain the posts deleted by the administrators of platforms, indicating there might be fake information in those threads. Therefore, our final sample ensures a relatively informative and rational online discussion setting and can reveal the sequential patterns in user posts. The final sample consists of 73,709 users and approximately 1.4 million posts from 29,086 threads. On average, each thread has 48 posts.

Text Mining

Text mining techniques were adopted to measure information coherence across user posts. Considering that the users participate in online discussion for different purposes, such as information seeking and experience sharing, topic modeling was applied to discover the underlying topic of each thread.

Content and Sentiment Coherence

Information coherence was measured by two dimensions, i.e., content and sentiment. To measure content coherence, we applied the Jaccard coefficient to calculate the lexical similarity between each pair of adjacent posts along the sequence of user posts, and aggregated the similarities to the thread level by taking an average. Previous research has shown that the Jaccard similarity coefficient has better performance for analyzing content (Tumuluru et al. 2012). We used the Jieba toolkit for word segmentation. We adopted the sentiment dictionaries in Chinese from the Linguistic Inquiry Word Count (LIWC) program (Pennebaker et al. 2007) for calculating sentiment coherence. We retrieved the positive and negative words of each post and constructed a vector consisting of three components: percentage of positive, negative, and neutral words. Sentiment coherence was measured by averaging the Cosine similarities between each pair of adjacent posts based on the sequence of user posts in each thread (Bhattacharjee et al. 2015). Both coherence measures take values between 0 and 1, and a higher value indicates a greater degree of information coherence in the thread.

Topic Discovery

We inferred the topic of each thread based on its initial post, considering that the initiators create threads for multiple reasons. Previous research has shown that the first post always contains information about the topic that users are interested in (Liu et al. 2010), leading to subsequent participation in a thread discussion.

We preprocessed the initial posts of the 29,086 threads with Jieba toolkit for word segmentation and stop word removal. The resulting corpus of the initial posts is transformed into a vector space of 147,068 words that comprise our dictionary. We then implemented latent Dirichlet allocation (LDA) (Blei et al. 2003) on our corpus to generate topics from words. The number of underlying topics was determined by model validation ranging from 2 to 15 topics. The model with 2 underlying topics was finally chosen because it produced the lowest perplexity. Based on the probabilistic distributions of words generated by the LDA model, we labeled the topic of each thread. Table 1 summarizes the most frequently used words belonging to each topic and the examples of initial posts which we have translated from Chinese.

Table 1. Keywords and Examples of Each Topic

Topic	Key words	Examples
Topic 1: Information acquisition	Oil Consumption, Automobile Brand, Space, Problem, Function, Compare	1. I want to ask those who already purchased car A. What is the oil consumption? Comparing to car B, which one is better?
		2. My tire is broken. Can this problem be fixed? How much it will cost me?
Topic 2: Experience sharing	Friend, Activities, Together, Life, Arrive, Place	1. My friend and his family came to my city during the weekend. I drive them around the city. We had a lot of fun together.
		2. It is snowing today! I am so lucky to have you by my side. Life is short but my love for you will never die.

In general, the discussion threads in our research context cover two topics: 1) information acquisition, driven by users' desire to search for information about products; 2) experience sharing, rising from users' tendency to share feelings related to product usage experience. Our result shows that most threads (21,252) serve for the information acquisition function, while 7,834 threads are related to experience sharing, reflecting that the platforms are mainly used for knowledge or information exchange about automobiles.

Stage 1: Coherence Patterns and Dependency

Exponential Smoothing

We used exponential smoothing to fit and predict the coherence patterns in each thread. Exponential smoothing is a method for data forecasting and is widely used in Finance and Operational Research (Kalekar 2004). The exponential smoother generates a forecast at time *t*, which is a weighted average of the current observation and the fitted value of previous observations. The formula is shown as follows.

$$\widehat{X_t} = \alpha X_t + (1-\alpha)\,\widehat{X_{t-1}} = \alpha X_t + \alpha(1-\alpha)\,X_{t-1} + \alpha(1-\alpha)^2\,X_{t-2} + \cdots$$

We applied the simple exponential smoothing by treating the Jaccard/Cosine similarity measures between each pair of adjacent posts as a series of information, along the sequence of posts. The smoothing factor α, ranging from 0 to 1, indicates how much weight the forecasted value puts on the current information. $1-\alpha$ is thus viewed as the *dependency* on the previous information. A higher degree of dependency indicates that the similarities of latter posts rely more on the patterns of earlier posts in the thread. We fitted the patterns of content and sentiment coherence respectively. For each thread, we split the data into training, validation, and testing set. We applied walk forward approach to have overfitting control and find the best estimate of the smoothing factor. Table 2 reports the average error over threads, indicating good fitting and prediction performance.

Table 2. Average Error over Threads of Validation and Testing Set

	Content Coherence		Sentiment Coherence	
	Validation	Testing	Validation	Testing
RMSE	0.034	0.035	0.116	0.117
RMSLE	0.031	0.032	0.074	0.075

Note: RMSE means root mean square error; RMSLE means root mean squared logarithmic error.

Summary Statistics and Results

Specifically, the mean of content coherence over threads is 0.12, which is lower than the mean of sentiment coherence (0.91, also see the results in Table 4). It suggests that the sentiment is overwhelmingly coherent while the content is a bit diversified in the threads in our research context. This is probably because users tend to show emotional support by following previous posts but provide different content to add new information. Another reason could be that we have included the neutral dimension in the calculation of sentiment coherence. The high level of sentiment coherence might result from people's preference to use neutral words in automobile discussion contexts. Regarding the dependency, i.e., $1-\alpha$, the dependencies of both content and sentiment are high (0.86 and 0.76), indicating that users rely much on the similarities of previous posts when contributing new ones to the discussion.

Table 3. Tests on the Differences of Coherence and Dependency between Topics

	Information Acquisition	Experience Sharing	T-test[1]	P-value
	Mean (N1=21,252)	Mean (N2=7,834)		
Content Coherence	0.12	0.11	-12.48	<0.001
Sentiment Coherence	0.94	0.86	-64.53	<0.001
Content Dependency	0.84	0.90	17.91	<0.001
Sentiment Dependency	0.73	0.82	20.90	<0.001

Note[1]: we also applied Wilcoxon signed-rank test without normal assumption. Results are consistent.

We further compared the information coherence and the dependency of threads between the two underlying topics, i.e., information acquisition and experience sharing. Table 3 shows significant differences. In particular, the discussion related to information acquisition has a higher level of content and sentiment coherence, compared to the coherence in experience sharing. It suggests that users are more concentrated on the content when discussing automobile features or problems. Meanwhile, their replies were relatively consistent in terms of sentiment (mostly neutral). In contrast, users rely more on the previous posts in the threads of experience sharing. The dependency of the threads about experience sharing is higher than that of information acquisition threads. This is probably because users care more about the patterns of previous posts in the threads where others share personal experience. As a response, they add posts by referring to the similarities of the previous discussions and show emotional support. However, users tend to keep unique and express their own opinions professionally in a problem discussion, leading to lower dependency on the previous posts in the threads of information acquisition.

Stage 2: Effects of Information Coherence

Stage 1 results demonstrate that there are inherent dependencies along the sequence of user posts. However, it is still not clear how the information coherence, aggregated by the similarities of posts, will take effect on overall user participation in the discussion. In this stage of analysis, we evaluate the effects of content and sentiment coherence on the online discussion at the aggregated thread level. Moreover, leveraging the dependency results from the stage 1 analysis, we further study the effects of information coherence, contingent on the dependency and the topics.

Variables and Model

We use *CONC* to represent content coherence and *SENC* to denote sentiment coherence. Two variables are considered to moderate the relationships between information coherence and online discussion. The first one is the dependency. We argue that the more users rely on previous posts, the more likely they would figure out the coherent patterns of previous information, thus leading to a concern of information redundancy. Leveraging the results in stage 1, we derived the dependency of content (*COND*) and of sentiment (*SEND*) in each thread. The second moderator is the topic. Users read online posts for different purposes. Upon the topic of each thread, users' expectations on the information coherence from previous posts may be different, result in different participation behaviors in the online discussion. Based on the topic discovery results, we coded a dummy variable (TOPIC) to indicate the topic category and the variable is equal to 1 if the topic of the thread is about information acquisition.

We controlled for a couple of factors related to thread and user characteristics. We included the number of views as a proxy for the overall popularity of a thread. We also controlled the number of days during which a thread is active within our observation window. A dummy variable was incorporated to control the difference of user base from the two platforms. In addition, we considered the characteristics of the initial post of each thread. We controlled the length of the initial post. We also incorporated the participation level of the user who created the thread and initiated the post. The variable was measured by the initiator's total number of posts on the platforms during our observation window. Moreover, we included time dummies of the initial post to capture unobserved events which might have happened every month.

Our dependent variable was measured as the total number of posts that a thread has received (*POSTS*). Table 4 summarizes the descriptive statistics for the main variables and their correlations. We calculated the variance inflations (VIFs) of the independent variables. All VIFs are lower than 3, suggesting that multicollinearity is not an issue in our analysis (Marquaridt 1970). Since our dependent variable is count data with overdispersion ($p<0.01$ by the likelihood-ratio test), negative binomial regression was applied to estimate the parameters. We standardized the main independent variables except the dummy one.

Table 4. Descriptive Statistics and Correlations (N =29,086)

	Variables	Mean	SD	Correlation					
				1	2	3	4	5	6
1	POSTS	47.51	60.32	1					
2	CONC	0.12	0.05	-0.04	1				
3	SENC	0.91	0.1	-0.42	0.08	1			
4	COND	0.86	0.31	0.18	-0.06	-0.09	1		
5	SEND	0.76	0.37	0.17	0	-0.14	0.11	1	
6	TOPIC	0.73	0.44	-0.33	0.07	0.35	-0.09	-0.11	1

Note: Indices 1–6 in the columns for correlation results represent the 6 variables.

Results

Table 5 shows the results. We estimated the main effects of information coherence, dependency, and topic in Model 1, and extended the model by adding interaction terms in Model 2 and 3. Model 4 is the full model.

Table 5. Results of the Effects of Information Coherence

	Model 1	Model 2	Model 3	Model 4
CONC	0.05 (3.0e-03)***	0.04 (3.1e-03)***	0.07 (6.1e-03)***	0.06 (6.1e-03)***
SENC	-0.11 (3.6e-03)***	-0.11 (3.6e-03)***	-0.11 (6.1e-03)***	-0.11 (6.1e-03)***
COND	0.06 (3.2e-03)***	0.06 (3.2e-03)***	0.06 (3.2e-03)***	0.06 (3.2e-03)***
SEND	0.03 (3.1e-03)***	0.04 (3.2e-03)***	0.03 (3.1e-03)***	0.04 (3.2e-03)***
TOPIC	-0.24 (7.2e-03)***	-0.24 (7.2e-03)***	-0.24(7.9e-03)***	-0.24 (7.9e-03)***
CONC*COND		-0.01 (2.5e-03)***		-0.01 (2.5e-03)***
SENC*SEND		-0.04 (3.3e-03)***		-0.04 (3.3e-03)***
CONC*TOPIC			-0.03 (7.0e-03)***	-0.03 (7.0e-03)***
SENC*TOPIC			0.01 (7.0e-03)	0.00 (7.0e-03)
Intercept	-1.52 (3.2e-02)***	-1.51 (3.2e-02)***	-1.53 (3.3e-02)***	-1.51 (3.3e-02)***
Controls	Included	Included	Included	Included
Pseudo R^2	0.84	0.85	0.84	0.85
-2*Log-likelihood	228900.19	228742.58	228886.40	228730.50
Notes: Standard errors in parentheses. *p<.10; **p < .05; ***p<.01.				

As shown in Model 1, content coherence and sentiment coherence have significant effects on online discussion. The number of posts in threads will increase if content coherence is high, while decrease with a higher sentiment coherence. The results suggest that users take a different view on the coherence of content and sentiment. Higher content coherence from previous posts reflects an informative and relevant discussion context, thus attracting more information exchange. However, higher sentiment coherence indicates consistently positive, neutral, or negative opinions in the previous discussion, which might enhance the feeling of information redundancy. As a result, users lose interest to further participate. The effects of content and sentiment dependency are both positive and significant. Higher dependency suggests a stronger thread cohesion. Therefore, users will contribute to the discussion, no matter the previous information is consistent or not. However, the total posts will be lower if the topic is for information acquisition. It suggests that people like to participate in the topic of experience sharing, which can help them to regulate emotional needs. Another reason might be that the threads about information acquisition are much more (approximately three times) than the threads of experiences sharing, result in more posts on the minority topic.

Model 2 evaluates the moderating role of dependency. The results show significantly negative moderating effects of dependency on the relationships between content/sentiment coherence and online discussion. When users' dependency on previous information is higher, it will demotivate their discussion because users are more likely to figure out the consistent content and sentiment from previous posts, which increase their concerns of information redundancy. We evaluated the moderating effect of the topic in Model 3. The coefficient of the interaction between content coherence and topic is significantly negative, suggesting that the positive impact of content coherence on online discussion will be attenuated when the threads are about information acquisition. The result is counterintuitive at a first glance but brings to light users' participation behaviors in the online discussion community. When users' purpose is to seek information, coherent content usually reflects a relevant and informative discussion and indicates that the problems might have been well resolved. In this case, users are less willing to extend the discussion by new posts. This finding also endorses users' desire to add meaningful information and make contributions to the community in digital collaboration. Model 4 shows consistent results. Our findings are summarized in Table 6.

Table 6. Summary on Preliminary Findings

Relationship	IV: Information Coherence	DV: Online Discussion
Main Effect	Content	Significantly positive
	Sentiment	Significantly negative

Moderation 1: Dependency	Content	Significantly lower the effect of content coherence under high dependency
	Sentiment	Significantly lower the effect of sentiment coherence under high dependency
Moderation 2: Topic	Content	Significantly lower the effect of content coherence under information acquisition (vs. experience sharing)
	Sentiment	Non-significant

Conclusion and Future Plan

In this paper, we investigate the influence of information coherence in the online discussion with a focus on the sequential patterns of user posts. Using the data from two popular online automobile discussion platforms in China, we empirically fit the patterns of coherence along the sequence of user posts and estimate the dependency by a data forecasting method. Moreover, we evaluate the effect of information coherence on online discussion and the moderating roles of dependency and topics. We measure the textual information by two dimensions, content and sentiment, to capture the content relevance and the sentiment consistency among the user posts. Our preliminary findings suggest that users are motivated to participate in the discussion when content is coherent, while their motivation will decrease with higher coherence in sentiment. Moreover, the findings reveal the synergetic relationships among information coherence, dependency, and topics in the online discussion.

Our study contributes to the existing literature by investigating the role of information coherence in digital collaboration. The empirical evidence sheds light on the heterogeneity of the coherence influence, which depends on a variety of factors, including dependency and topics. Our findings, therefore, provide valuable implications for practice to facilitate online discussion. Content relevance is shown to be more important than sentiment consistency to attract online posts. However, while a content relevant discussion may help resolve users' problems quickly, there is a cost to encourage further discussion. Instead, sharing experience might create fun and thus engage users better. Our study brings insights into the strategy designs for online discussion platforms and business managers to improve user engagement and motivates a comprehensive study to understand the role of the interventions on information coherence and discussion topics in digital collaboration.

This research work is still in progress. We plan to further enhance this study in two aspects. First, we will enrich the theoretical framework by better integrating the theories on online discussion context, information coherence, information processing, and the conceptualization of content and sentiment factors. Second, we will strengthen our empirical analysis by trying different measures of content and sentiment coherence and adding more robustness checks, such us using instrumental variables to address the reverse causality problem. We also plan to extend our dependent variable and include more controls for thread and user characteristics.

Acknowledgement

This study is supported by the ECS Fund from the Research Grants Council of Hong Kong [25508819].

References

Bhattacharjee, S., Das, A., Bhattacharya, U., Parui, S. K., and Roy, S. 2015. "Sentiment Analysis Using Cosine Similarity Measure," *2015 IEEE 2nd International Conference on Recent Trends in Information Systems (ReTIS): IEEE*, pp. 27-32.

Blei, D. M., Ng, A. Y., and Jordan, M. I. 2003. "Latent Dirichlet Allocation," *Journal of machine Learning research* (3: Jan), pp. 993-1022.

Chen, Z., and Berger, J. 2013. "When, Why, and How Controversy Causes Conversation," *Journal of Consumer Research* (40:3), pp. 580-593.

Cheung, C. M., Lee, M. K., and Rabjohn, N. 2008. "The Impact of Electronic Word-of-Mouth: The Adoption of Online Opinions in Online Customer Communities," *Internet Research: Electronic Networking Applications and Policy* (18:3), pp. 229-247.

Chevalier, J. A., and Mayzlin, D. 2006. "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of marketing research* (43:3), pp. 345-354.
Das, S. R., and Chen, M. Y. 2007. "Yahoo! For Amazon: Sentiment Extraction from Small Talk on the Web," *Management science* (53:9), pp. 1375-1388.
De Beaugrande, R.-A., and Dressler, W. U. 1981. *Introduction to Text Linguistics*. Longman London.
Dellarocas, C., Zhang, X. M., and Awad, N. F. 2007. "Exploring the Value of Online Product Reviews in Forecasting Sales: The Case of Motion Pictures," *Journal of Interactive marketing* (21:4), pp. 23-45.
Dewan, S., Ho, Y.-J., and Ramaprasad, J. 2017. "Popularity or Proximity: Characterizing the Nature of Social Influence in an Online Music Community," *Information Systems Research* (28:1), pp. 117-136.
Ehrlich, D., Guttman, I., Schönbach, P., and Mills, J. 1957. "Postdecision Exposure to Relevant Information," *The journal of abnormal and social psychology* (54:1), p. 98.
Feather, N. T. 1962. "Cigarette Smoking and Lung Cancer: A Study of Cognitive Dissonance," *Australian journal of Psychology* (14:1), pp. 55-64
Geeraerts, D., and Cuyckens, H. 2007. *The Oxford Handbook of Cognitive Linguistics*. Oxford University Press.
Godes, D., and Silva, J. C. 2012. "Sequential and Temporal Dynamics of Online Opinion," *Marketing Science* (31:3), pp. 448-473.
Kalekar, P. S. 2004. "Time Series Forecasting Using Holt-Winters Exponential Smoothing," *Kanwal Rekhi School of Information Technology* (4329008:13), pp. 1-13.
Kane, G. C., Johnson, J., and Majchrzak, A. 2014. "Emergent Life Cycle: The Tension between Knowledge Change and Knowledge Retention in Open Online Coproduction Communities," *Management Science* (60:12), pp. 3026-3048.
Kleck, R. E., and Wheaton, J. 1967. "Dogmatism and Responses to Opinion-Consistent and Opinion-Inconsistent Information," *Journal of Personality and Social Psychology* (5:2), p. 249.
Liu, D., Percival, D., and Fienberg, S. E. 2010. "User Interest and Interaction Structure in Online Forums," *arXiv preprint arXiv:1009.1555*.
Ma, L., Sun, B., and Kekre, S. 2015. "The Squeaky Wheel Gets the Grease—an Empirical Analysis of Customer Voice and Firm Intervention on Twitter," *Marketing Science* (34:5), pp. 627-645.
Manchanda, P., Packard, G., and Pattabhiramaiah, A. 2015. "Social Dollars: The Economic Impact of Customer Participation in a Firm-Sponsored Online Customer Community," *Marketing Science* (34:3), pp. 367-387.
Marquaridt, D. W. 1970. "Generalized Inverses, Ridge Regression, Biased Linear Estimation, and Nonlinear Estimation," *Technometrics* (12:3), pp. 591-612.
Moe, W. W., and Trusov, M. 2011. "The Value of Social Dynamics in Online Product Ratings Forums," *Journal of Marketing Research* (48:3), pp. 444-456.
Pennebaker, J. W., Booth, R. J., and Francis, M. E. 2007. "Liwc2007: Linguistic Inquiry and Word Count," *Austin, Texas: liwc. net*.
Streufert, S. C. 1973. "Effects of Information Relevance on Decision Making in Complex Environments," *Memory & Cognition* (1:3), pp. 224-228.
Tumuluru, A. K., Lo, C.-k., and Wu, D. 2012. "Accuracy and Robustness in Measuring the Lexical Similarity of Semantic Role Fillers for Automatic Semantic Mt Evaluation," *Proceedings of the 26th Pacific Asia Conference on Language, Information, and Computation*, pp. 574-581.
Villarroel Ordenes, F., Ludwig, S., De Ruyter, K., Grewal, D., and Wetzels, M. 2017. "Unveiling What Is Written in the Stars: Analyzing Explicit, Implicit, and Discourse Patterns of Sentiment in Social Media," *Journal of Consumer Research* (43:6), pp. 875-894.
Wang, C., Zhang, X., and Hann, I.-H. 2018. "Socially Nudged: A Quasi-Experimental Study of Friends' Social Influence in Online Product Ratings," *Information Systems Research* (29:3), pp. 641-655.
Wang, Y., and Chaudhry, A. 2018. "When and How Managers' Responses to Online Reviews Affect Subsequent Reviews," *Journal of Marketing Research* (55:2), pp. 163-177.
Wasko, M. M., and Faraj, S. 2005. "Why Should I Share? Examining Social Capital and Knowledge Contribution in Electronic Networks of Practice," *MIS quarterly*, pp. 35-57.
Wyer Jr, R. S. 1970. "Information Redundancy, Inconsistency, and Novelty and Their Role in Impression Formation," *Journal of Experimental Social Psychology* (6:1), pp. 111-127.