Available online at www.sciencedirect.com
ScienceDirect
Journal of Interactive Marketing 28 (2014) 55 – 67
JOURNAL OF INTERACTIVE MARKETING
www.elsevier.com/locate/intmar

Thematic Discrepancy Analysis: A Method to Gain Insights into Lurkers and Test for Non-Response Bias

Scott A. Thompson [a,*] & James M. Loveland [b] & Paul W. Fombelle [c]

[a] Department of Marketing, Terry College of Business, The University of Georgia, Athens, GA 30602, USA
[b] Department of Marketing, HEC Montréal, 2953, Chemin de la Côte-Sainte-Catherine, Montréal, Quebec H3T 1C3, Canada
[c] Marketing Group, College of Business Administration, Northeastern University, 202 Hayden Hall, Boston, MA 02115, USA

Available online 6 August 2013

Abstract

Word of mouth (WOM), long recognized as a highly influential source of information, has taken on new importance with the proliferation of online WOM. Research in online environments has focused on individuals who actively participate in generating WOM. However, over 90% of those that read WOM are non-participants, commonly called “lurkers.” This paper develops and tests a thematic discrepancy analysis (TDA) approach that combines commonly available information on *Views* and *Replies* with content analysis to provide new insights into differences between WOM participants and lurkers. TDA provides managers with market-sensing information to identify hidden opportunities and threats, as well as to test for non-response bias. Given the lack of approaches to address non-response bias due to lurkers, TDA represents a significant contribution to research methodology. We demonstrate the efficacy of TDA by applying it to a large scale WOM dataset containing over 80,000 messages from a brand-specific online forum.

Keywords: Word of mouth; Research methodology; Non-response bias; Market sensing; Lurkers

Introduction

The rise of online forums and communities has dramatically increased the scope of word of mouth (WOM) marketing (Dwyer 2007; Kozinets et al. 2010), allowing consumers greater access to information from subject matter experts and other key influentials who impact new purchases (Rogers 2003). However, the value of these forums extends well beyond the ability to facilitate WOM communication—the forums are also important locales for brand admirers to communicate and share experiences, further enhancing the value and resonance of the brand (Keller 2003) in ways that are simply not possible in traditional, face-to-face settings.

Online forums provide detailed WOM data that were unattainable prior to the advent of the Internet. Using WOM data from forums, both managers and researchers can now see exactly what was said and when it was said, often time-stamped down to the second. In these forums, messages are organized into “threads” comprised of an original post or review, followed by replies submitted to the thread by participants. This organization allows managers, researchers, and community members alike to see who is replying to whom. Furthermore, online forums expose a wealth of related data, including the number of people replying to a thread and even the number of times a thread has been viewed by visitors.

These forums attract a broad array of consumers including brand aficionados, product category experts, and general enthusiasts, as well as individuals simply seeking information. Naturally, research on forums and Internet communities has focused on the activity of experts, key influentials, innovators, and enthusiasts, as those individuals are the ones who provide the bulk of the activity on these sites. Services such as Google Analytics (Google 2012) monitor general activity along key search terms, or unusual activity at a firm’s site, allowing managers to monitor discussions among active participants on previously identified topics or themes. In contrast, comparatively little research has examined the non-participants or “lurkers” who “hear” this WOM and silently consume the information (e.g., Mathwick 2002).

* Corresponding author.
E-mail addresses: sat@uga.edu (S.A. Thompson), james.loveland@hec.ca (J.M. Loveland), p.fombelle@neu.edu (P.W. Fombelle).

1094-9968/$ -see front matter
http://dx.doi.org/10.1016/j.intmar.2013.06.001

Furthermore, managers lack a means to identify the themes that are of interest to these lurkers, as other analytical approaches confine the purview of managers to only those themes which they have specifically designated. If managers cannot identify in advance what non-customers are interested in, then these monitoring services are excluding an important group of potential customers.

This represents an important gap in the research, as these lurkers typically make up 90% or more of the traffic in online forums and communities (Nielsen 2006) and often comprise the audience of interest for research and managers alike. This gap in the research is increasingly problematic when we consider the massive and rapidly growing volume of Internet communities and forums, and the fact that many of the conclusions being made about the power of these forums are based on potentially unrepresentative segments of the target population.

Being able to determine when and how the interests of lurkers diverge from those of active WOM participants would create managerial opportunities as well as address serious methodological issues. First, identifying areas that are of more interest to lurkers than to active participants would enable firms to behave proactively to emerging opportunities and threats, before competitors became aware of them. Specifically, firms would be able to identify opportunities based on a rise in interest in a particular product or feature among lurkers before the volume of posts and replies among active participants reflects this increased interest. Similarly, firms would be able to detect emerging problems, such as a rising interest in a new feature offered by a competitor, before these problems harmed sales, even if the active participants did not show a similar interest. Although competing firms have access to public forums as well, employing thematic discrepancy analysis (TDA) provides managers with a unique advantage in that it provides insights into those consumers who do not actively participate, and thus whose interests are not evident from simply looking at forum participants. These forums also attract more views and activity than do corresponding brand affiliated sites. Thus, these forums allow for comparative insights that would not otherwise be available from just focusing on the firm's current customers. This also allows managers to identify competitive advantages based on issues and features that are important to prospective consumers who are seeking information, rather than just focusing on the vocal minority of active participants. Perhaps most importantly, the TDA approach allows managers to identify the themes of importance to the larger audience of lurkers, through the lurkers' own viewing activity. This provides a valuable alternative to managers deciding a priori what issues prospective customers may or may not be interested in.

Second, a method to identify differences in interests between lurkers and participants would be a boon to managers and researchers, as it would permit researchers to evaluate the generalizability of their results. Such a method would provide an ad hoc test for non-response bias. Finally, managers would be able to determine whether agitation over an issue among active participants reflected a broader problem among the vast majority of lurkers. This would assist managers in deciding when and whether to respond to issues while avoiding potentially costly overreactions to a vocal minority.

In response to these managerial challenges and opportunities, we develop and test a new method, called thematic discrepancy analysis (TDA), which analyzes changes in the ratios of WOM activity and WOM views in order to reveal which WOM themes lurkers are reading. This allows researchers and managers to identify any differences from or similarities to active participants. We then describe how the TDA approach can be used in a discovery or confirmatory capacity. TDA identifies potential opportunities and threats that managers would miss if only the active WOM participants are considered by identifying discrepancies between the interests of participants and lurkers. Moreover, TDA can be used to assess differences between participants and lurkers on specific topics of managerial interest or concern. This approach provides researchers with a new method to empirically test for non-response bias, allowing them to evaluate the generalizability of conclusions and rule out alternative hypotheses.

We demonstrate the efficacy of the TDA technique using online WOM data, consisting of over 80,000 messages from members of a brand-specific Internet forum. Using these data, we conduct a TDA using scatterplots and structural equation modeling (SEM) combined with a form of content analysis. The results reveal that lurkers are disproportionately interested in certain themes, while not in others. Our results also demonstrate how managers can employ TDA to actively monitor their online forums and communities, not only as a means of generating Internet marketing metrics, but also as a means to monitor changes in interest in critical pre-determined topics.

This paper is organized as follows: first, we review and integrate the literatures on non-response bias, and Internet forums and communities, and we review some of the basic metrics that have been used to monitor activity in these forums. Then, we introduce a TDA approach that addresses challenges identified in the extant literature and discuss the two different variants of the method. We then demonstrate our TDA approach using large scale WOM data. We show how the method can be used to reveal new, actionable insights for managers. Furthermore, we demonstrate how TDA can be employed to test for non-response bias due to differences between WOM participants and lurkers. Finally, we discuss the theoretical and managerial implications of TDA, as well as limitations and research opportunities.

Online Word of Mouth in Marketing

The rise of the Internet has facilitated an explosion of WOM activity. In response, firms and researchers alike have been scrambling to harness its potential as both a conduit for WOM and a source of data on WOM behavior. Online WOM data have been widely used in the literature (c.f., Kozinets 2002; Kozinets et al. 2010) to examine topics such as the impact of WOM recommendations and reviews (Liu 2006), brand community involvement (Muñiz and O'Guinn 2001), and product adoption (Algesheimer and Dholakia 2006; Thompson and Sinha 2008). Of course, consumers also use online environments to express their dissatisfaction with the brand or its products (e.g., Grégoire, Laufer, and Tripp 2010; Hennig-Thurau et al. 2004). Finally, online forums are increasingly being used to conduct targeted advertising campaigns through banner ads (e.g., Morrissey 2010; Sheng 2007; Tran 2000) and indirectly through paid Google searches. These environments provide a ready-made audience for product categories or even specific brands, allowing for highly targeted marketing. Additionally, many of these sites attract a very high number of views from visitors, ranging from hundreds, to hundreds of thousands, per day. Thanks to the nature of the Internet, the size of the audience is readily measurable and the number of views easily tracked.

Targeting decisions rely on the content posted on these sites by active members (Sheng 2007). Thus, online ads for particular brands often target online environments where a large number of that brand's fans participate. However, over 90% of visitors who view this content are "lurkers," consumers who visit these sites and read the WOM discussions but do not actively participate in them (Nielsen 2006; Mathwick 2002). This represents a serious challenge for managers hoping to take advantage of these market-sensing and targeted advertising opportunities. Assuming that the small number of active WOM participants is representative of the vast majority of non-participating lurkers is risky. WOM participants might be more interested in some features or topics than non-participants, leading managers to make erroneous conclusions about the interests of the target market. Similarly, WOM participants may also be under-emphasizing other aspects that are of greater interest to the silent majority, leading to missed opportunities.

Online WOM Research and Non-response Bias

Differences between lurkers and participants also represent a problem for WOM researchers and managerial decision makers. If WOM participants differ in their interests from non-participants, conclusions based on the participants may be compromised by selection effects in the form of non-response bias (c.f., Cook and Campbell 1979). Non-response bias occurs when there is some intervening variable which differentially affects participation in a study in such a way that the participants do not represent non-participants. Researchers attempting to allay concerns of non-response bias have relied on making comparisons between responders and non-responders on known basic characteristics of the sample, such as age, race, and firm size, and by comparing early versus late responders (Armstrong and Overton 1977). However, this approach assumes that there are not any mediating reasons some would choose not to participate in said research.

Within the marketing literature, research suggests that such differences between WOM participants and non-participants are common. This is especially alarming when one considers that much of the online WOM behavior of interest to marketers is already rather endogenous (Godes and Mayzlin 2004), as only those consumers who own a product or have experienced a service could discuss it online. In addition, many online environments such as forums require users to provide information including email addresses in order to receive an account that permits them to post WOM messages. Joining the PlayStation Network, for instance, requires the user to provide his/her name, address, e-mail, and credit card information to receive an account. In turn, an account is required to contribute to WOM in these forums. Thus, WOM participants are likely to differ from lurkers in their willingness to share such information.

Research has also shown that differences in participation between types of online environments impact important marketing outcomes. For instance, Thompson and Sinha (2008), in their study of participants of both brand-specific and product category forums, found that differences in participation rates across brand forums significantly affected how quickly the different forum members adopted the newest computer processor. However, the degree to which consumers participated in general forums did not have the same effects on new product adoption. These findings suggest that the impact of online WOM activity differs across settings. Similarly, Hickman and Ward (2007) noted that many WOM participants of communities actively engage in mocking the product failures of other communities. However, if participants self-selected to participate based on their love of a brand, the interest in such WOM may not reflect the interests of non-participants. In short, there is ample evidence to suggest that non-response bias represents a threat to conclusions about interests or concerns based on the content of online WOM.

Assessing Non-response Bias Due to Lurkers

Yet simply because WOM participants differ from non-participants does not mean that a given study's conclusions suffer from non-response bias. Just as they differ in some interests, participants and non-participants will likely be similar in their interests in other areas. The problem facing marketing researchers, then, is to identify whether WOM participants differ from lurkers on the specific WOM topics they are studying and attempting to draw conclusions about.

Conventional approaches for assessing non-response bias rely on comparing early to late responders (e.g., Armstrong and Overton 1977). This approach relies on empirical research showing that individuals who take longer to respond to a mail survey are more similar to non-respondents than they are to those participants who responded sooner to the survey. This approach has been extended to evaluate online survey data for non-response bias despite significant concerns expressed by researchers (c.f., Blair and Zinkhan 2006). However, this approach is inadequate for the task of testing online WOM for non-response bias for several reasons. First, online WOM is a continuous process, with no discrete starting point such as the mailing of a survey. Thus, it is difficult to distinguish "early" from "late" participants. Second, online WOM participants are not responding to a study request when they engage in WOM behavior. Thus, the variety of motives separating participants from non-participants is likely to differ from those separating study respondents to non-respondents. As a result, making inferences about the interests of lurkers based on how recently WOM participants posted WOM content does not provide a means to test for non-response bias. While some researchers (e.g., Dillman 2000) have developed processes for assuring high-quality Internet and mail-based surveys, this approach still requires identification of non-responders (lurkers), who are, by definition, not identifying themselves. In sum, managers and researchers alike lack any method to address the serious threat to generalizability inherent to most online settings due to the lack of information on lurkers.

Thematic Discrepancy Analysis

A key challenge facing managers and researchers is the need for a method that provides insights into whether and how the interests of the silent majority of lurkers differ systematically from active WOM participants. In response to this clear need, we developed the TDA approach which not only allows marketers to identify the topics that appeal to lurkers, but also provides researchers with a means to evaluate non-response bias. By definition, individual level information is unavailable on lurkers, and a lurker in one discussion may or may not be a participant elsewhere. In fact, many individuals likely engage in some type of lurking behavior, and so this behavior would represent a constant across themes. Nonetheless, if participants in one area lurk in another without posting, the result could be a discrepancy between the participants in the second area and these lurkers. Therefore, TDA uses information drawn from message threads in online WOM environments, *which represent specific WOM discussions where some consumers are participating while others are lurking*. TDA achieves this using a three-stage process that utilizes commonly available, but overlooked, information in online forums combined with content analysis. Fig. 1 illustrates the process that comprises TDA.

The first stage involves collecting and analyzing publicly available information on the number of times a message thread has been viewed as well as the number of times it has been replied to. Thread view counts and reply counts are widely tracked and reported by the software platforms used to host online forums. Indeed, this information is automatically collected and displayed by the leading platforms, including the market leader vBulletin (W3Techs 2011). View counts are publicly available information which reflect the number of times someone, whether a participant or not, has viewed a particular discussion thread, message, or review. Thus, view counts provide a measure of the impact that the WOM in the thread or messages has had in general, with higher number of views indicating a greater *WOM impact*. Reply counts, which are publicly available information, reflect the number of times a participant has contributed a message to a thread or commented on a review. Higher numbers of replies indicate a higher level of *WOM activity*.

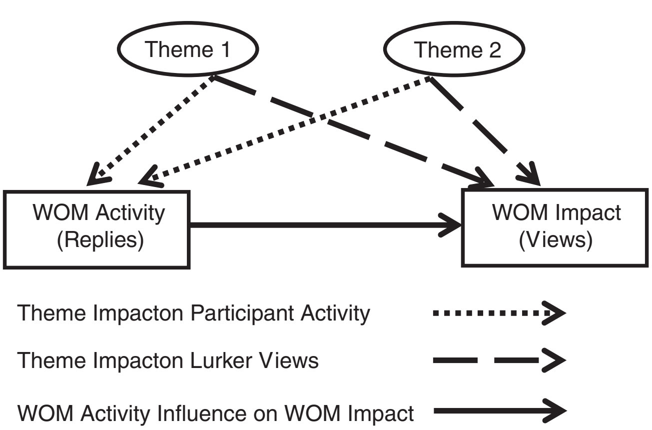

Fig. 1. Thematic discrepancy analysis.

In the second stage of TDA, the text content of the WOM messages comprising the threads is collected. This WOM content is then content analyzed to identify topics or themes in the vast amount of WOM. The result of this stage is a total list of themes contained in all of the WOM as well as measures indicating the presence or absence of those themes in the WOM constituting each thread.

In the third and final stages of TDA, themes that show a discrepancy in the relationship between WOM activity and WOM impact are identified. In this study, we used a combination of scatterplots, structural equation modeling, and mediation analysis to identify these discrepancies, but other traditional statistical methods are also appropriate. TDA is predicated on the idea that, in the absence of lurkers, *WOM impact* will be driven by *WOM activity*, with topics of interest to WOM participants leading to an increase in replies and hence views as shown by the solid line in Fig. 1. However, for those threads or reviews containing information or themes of interest to lurkers, there will be a rise in the number of views without an associated rise in replies. Thus, TDA identifies themes of interest to lurkers based on a theme's discrepant impact on views, above and beyond impact on replies (the dashed line in Fig. 1). Specifically, the test for mediation of views by replies assesses whether or not the replies are actually driving views, or if the views are being driven by discrepant interest in a thread from lurkers. Detecting these discrepancies allows TDA to identify those themes which are of particular interest to the population of lurkers relative to WOM participants, as these will be the themes whose WOM impact will be mediated by WOM activity.

Discovery versus Confirmatory TDA

Depending on the nature of the managerial or research problem, TDA can be employed in either a discovery oriented or confirmatory manner. Discovery oriented TDA involves identifying the themes that account for discrepancies without a preconceived set of topics in mind. In this case, the goal is to identify themes of particular interest to lurkers. This allows managers to discover opportunities and potential problems that the marketer might not be aware of. Furthermore, a discovery oriented TDA can reveal marketing and advertising messages that the lurkers will be more receptive to than the active participants.

TDA can also be used with a confirmatory orientation, to compare multiple online contexts to identify those communities, forums, and threads where lurkers show a discrepant interest in a particular theme—e.g., WOM comparing a firm's brand with competitors. Banner ads can then be placed which favorably compare the firm's brand with those competitors. In addition, conducting confirmatory TDAs over time allows managers to monitor a community to confirm the presence of a pre-determined topic, enabling them to detect increases associated with problematic themes such as complaints or requests for help.

TDA and Content Analysis Techniques

One of TDA's strengths lies in the fact that it is not dependent on the form of content analysis used. The themes can be based on pre-determined categories of specific words or terms (such as a "complaint" theme which contains "problem," "dissatisfied," "failed," etc.), pairings of terms (such as a "Honda Performance" theme comprised of "Accord + speed," "Civic + mileage," etc.), or themes identified inductively by non-dictionary approaches. Furthermore, the content analysis can be conducted by hand, or automated using computer software. This flexibility allows firms to employ TDA with existing software tools and data packages that they may already be using to monitor WOM.

However, the type of TDA being conducted has implications for whether a dictionary based technique or a non-dictionary based technique is more appropriate. Dictionary based content analysis techniques involve using dictionary files, containing pre-existing sets of words or terms, to search and code bodies of text (e.g., Pennebaker et al. 2007; Pennebaker, Mehl, and Niederhoffer 2003). In addition, dictionary based approaches may be prone to failure in online forums where community-invented terms (jargon, slang, or abbreviations) are used between members. Thus, these approaches can only detect discrepancies related to topics identified in advance. Given this, dictionary based techniques are most appropriate when conducting confirmatory TDA.

On the other hand, non-dictionary based content analysis techniques involve any content analysis that does not rely on a pre-determined set of terms (e.g., Alterman and Bookman 1990; Carley 1997; Humphrey 1999; Corman et al. 2002). Non-dictionary based techniques excel at identifying themes or topics that might be overlooked by managers. As such, these techniques are preferable for conducting discovery oriented TDA as they do not require a priori assumptions. However, non-dictionary based content analysis is also suitable for confirmatory TDAs where the researcher is interested in a particular topic but is unsure which terms may be relevant. In this case, a non-dictionary based technique could be used to identify themes and the associated nomological networks. The researcher would then focus on evaluating the discrepancies associated with the themes that match the topics previously determined to be of interest.

Thematic Discrepancy Analysis in Practice

To demonstrate the application of TDA, we conducted a discovery oriented TDA followed by a confirmation oriented TDA using online WOM data from a high traffic message board dedicated to Intel microprocessors. This context was chosen for two reasons. First, the forum is hosted by one of the highest traffic computer enthusiast websites on the Internet. Thus, it is managerially relevant both as a source of market-sensing information as well as an advertising venue. Second, data from computer contexts in general and this context in particular have been used in previously published academic research (e.g., Muñiz and Schau 2005; Thompson and Sinha 2008). Thus, it represents a context in which evaluating non-response bias is an important consideration for existing and future research.

The objective of the discovery oriented TDA is to discover themes that are being viewed disproportionately by lurkers in order to better target banner ads on the site and to refine marketing messages. In the confirmatory TDA, we first study the impact of problems with a complementary product (the Windows Operating System, or OS) on WOM about the focal product (Intel processors) to test for non-response bias. Complementary products represent a topic of growing interest to researchers and managers (e.g., Aribarg and Foutz 2009). Thus, this represents an application of TDA to a salient research problem. Specifically, we consider a situation in which researchers are concerned that active WOM participants may be disproportionately interested in complementary products, and thus the WOM data overstate, relative to non-participants, the role of complementary products. In addition, we test whether there exists a bias in a second forum topic identified as a theme in the discovery oriented TDA: the competitive comparisons of AMD vs. Intel. Again, we are seeking to see if WOM participants differ disproportionally on their interest in competitive comparisons.

We begin by describing the data used to conduct the two TDAs and discussing the non-dictionary based content analysis used to identify themes in the WOM. We then plot the number of views and replies for each thread to identify which themes show the greatest discrepancies in views. Next, we employ structural equation modeling to detect discrepancies in the identified themes. Finally, we show how these analyses are used to perform both the discovery and the confirmatory TDAs.

Data

For the analyses, we chose a WOM dataset comprised of the threads collected from a message board dedicated to Intel microprocessors. The final dataset contains 5806 threads spanning a 36 month period from January 1, 2004 to February 16, 2007. This time frame was chosen in part because this reflects the typical product cycle of three years in this industry. Moreover, TDA provides a unique approach for detecting persistent differences around which managers might need to orient longer-term strategy. Still, this approach could also be used over a much shorter time frame depending on the research needs of the manager. These threads contain 80,574 WOM messages posted by 6,548 unique user accounts. For each thread, counts of replies, views, and the actual message content were collected on a daily basis. The *Replies* variable indicates the number of messages posted in response to the original message and thus is a measure of WOM activity. *Views* is a measure of the number of times individuals have read the messages in the thread. While an individual has to be registered on the website in order to post a message, anyone can view the messages. Therefore, *Views* reflects how many members and non-members have read the WOM and is thus a measure of WOM impact.

Content Analysis

To identify themes within the threads, a non-dictionary based content analysis technique was chosen. Non-dictionary based approaches are appropriate from both variants of TDA and provide a more rigorous test since they are free of the assumptions inherent in researcher selected dictionaries. First, each thread was content analyzed using centering resonance analysis (CRA), a representational content analysis technique based on centering theory (Corman et al. 2002). CRA treats words as objects connected into networks which produce meaning. Influence scores ranging from 0 to 1 are calculated for each word in the network based on its centrality, or the degree to which the word serves to connect other words together. Since CRA generates representations of texts without relying on dictionaries, semantic networks, or ontologies, it is well suited to analyzing texts which contain unique or technical terminology that evolves over time. As such, this type of content analysis is ideal for analyzing the messages from technical enthusiast forums such as the one examined in this study. This approach provides a very simple and cost-effective approach to capture differences in themes, it does not require the researcher to build a dictionary unique to the research setting, and it does not overly reduce the complexity of the data.

In the next step, an exploratory factor analysis using principal axis factoring with a varimax rotation was performed. Since the dataset includes over 80,000 messages, there is a large number of themes present. In fact, our initial EFA identified 99 factors with eigenvalues above 1.0. Since a central goal of a TDA is to examine the most important themes to lurkers and participants, we focused on the six factors which had an eigenvalue above 2.0, as these factors parsimoniously accounted for variance, exceeded the Kaiser criterion of 1.0, and allowed for a clear demonstration of the TDA approach. However, it is important to note that TDA could be performed using higher eigenvalues to reduce the number of factors further, or lower eigenvalues if consideration of a larger number of themes is valued over parsimony, depending on the specific managerial or research problem it is being employed to address.

The six factors represented themes involving: Thermal Issues (F1), Overclocking (F2), the construction of New Computer Systems (F3), comparisons of Intel and AMD processors (F4), Operating System (OS) issues (F5), and Memory Speed issues (F6). To produce reasonably parsimonious models, only words that demonstrated factor loadings equal to or greater than .20 were included in the models. The use of .20 represents a conservative cut-off point that ensured that the full range of topics receiving interest were included in the models.

The results of the EFA are reported in Table 1. While the exclusion of the remaining 93 factors which had very low factor loadings (<.20) allows for more parsimonious models, the limitations to this approach must be kept in mind. First, their exclusion reduces the potential fit of the resulting models. Second, the fact that some words load on more than one factor raises the prospect that the exclusion of factors might be problematic. To address this concern, we examined the correlation between the words associated with each factor. In the case of Intel vs. AMD (F4), ten words were originally identified by EFA. However, four of the words were correlated with one another, while the remainder had weak correlations ($p < .05$) with this group of four words. This pattern indicates that the other six words were elements of related themes that had been excluded due to having an eigenvalue of less than 2.0. As a result, the fourth theme is comprised of the four correlated words.

Table 1
Correlations between the items (words) and the top six factors (themes).

Factor 1: Thermal issues		Factor 2: Overclocking		Factor 3: New computers		Factor 4: AMD vs. Intel CPUs		Factor 5: O/S issues		Factor 6: Memory issues	
Items	Loading	Items	Ldg.	Items	Ldg.	Items	Ldg.	Items	Ldg.	Items	Ldg.
Fan	0.45	Stable	0.47	Card	0.35	AMD	0.32	Window	0.37	Fsb	0.36
Temps	0.42	Voltage	0.45	Ram	0.28	Intel	0.30	Xp	0.31	Speed	0.33
Heatsink	0.39	Vcore	0.37	Board	0.27	Performance	0.30	Drive	0.29	Memory	0.27
Load	0.37	Prime	0.32	Good	0.26	Game	0.28	Driver	0.25	Ram	0.24
Idle	0.37	Ortho	0.32	New	0.24			Problem	0.22	High	0.23
Hsf	0.33	High	0.27	Mobo	0.22			Os	0.21	Clock	0.22
Zalman	0.33	Hour	0.26	Sata	0.22			Raid	0.21		
Hot	0.32	Test	0.23	Video	0.21						
Stock	0.31	Overclock	0.22	Asu	0.21						
Cooler	0.31	Stock	0.22								
Case	0.31	Volt	0.22								
Temp	0.30	Setting	0.21								
Thermal	0.29	Ram	0.21								
Paste	0.28	Temp	0.21								
Degree	0.22										
Cooling	0.22										
Silver	0.21										

Note. Table includes the top six factors extracted using principal axis factoring (varimax rotation). Only those items with factor loadings equal to or greater than .20 are reported.

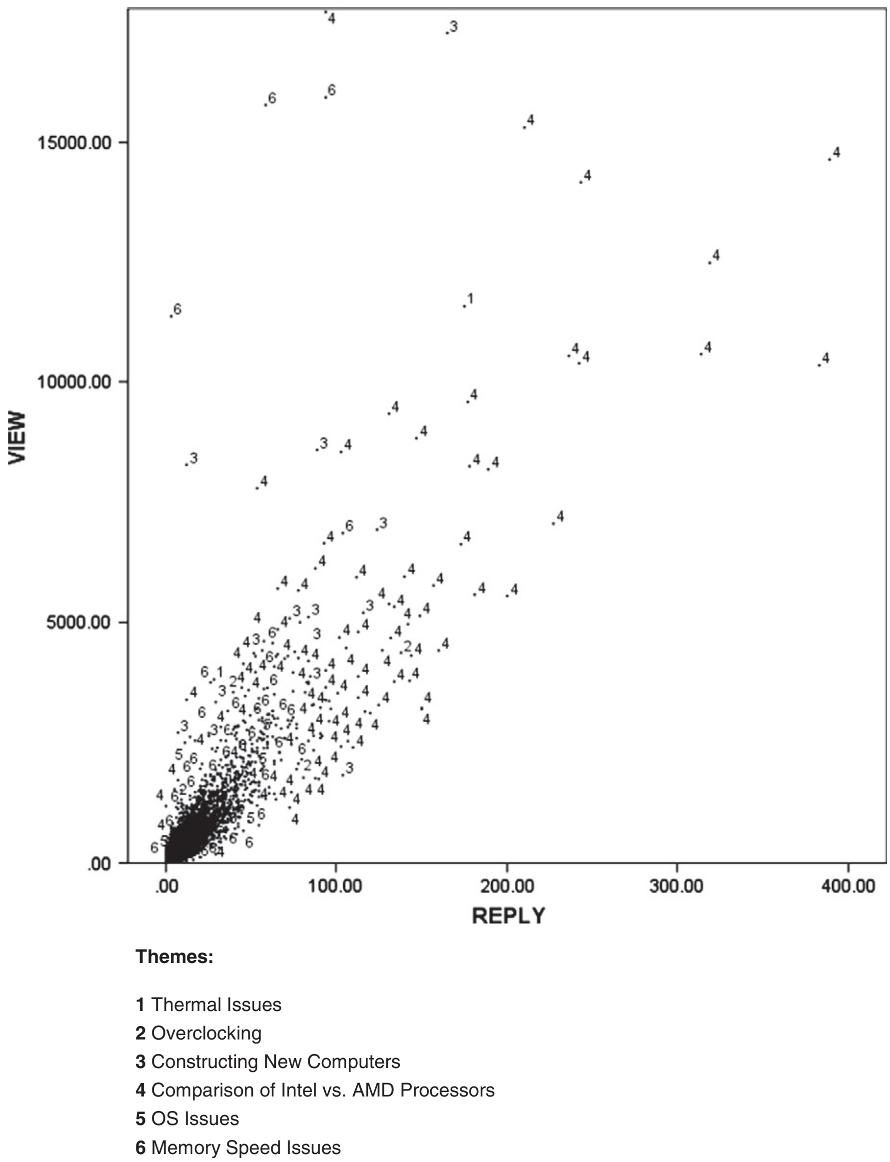

Fig. 2. Views and replies for each thread by theme.

Results for TDA Analysis

To detect discrepancies associated with the themes, the number of views and replies for threads associated with each theme is first plotted. Each thread is categorized based on the theme for which it has the highest factor loading. This provides a simple visual representation of the level of discrepancy associated with each thread, as well as the overall distribution of the identified themes. This plot serves as a quick way for managers to visually scan the data and to identify the themes that are garnering attention from both active participants and lurkers. The scatterplot using the data from the forum in question appears in Fig. 2.

The scale of views versus replies is notable. In comparison, the number of replies is very small relative to the number of views, indicating the presence of large discrepancies for some of the specific threads. At the same time, there are many threads clustered in a very narrow range, with a small number of both replies and views. Examining the themes that lie outside of this cluster, we also find that threads associated with the Intel vs. AMD (F4) theme overwhelmingly predominate. However, there are also some threads associated with other themes such as New Computer Systems (F3) and Memory Speed issues (F6). Based on this cursory examination, it seems possible that discussions of AMD versus Intel (F4) products show a discrepant level of views, relative to other themes.

While providing managers with a quick way to identify clearly discrepant themes, examining the scatterplot of the threads has limitations. Specifically, it does not indicate whether the apparent differences are statistically significant, especially for themes that are not as clearly discrepant. Thus, it is not clear whether there are any meaningful differences between the themes for Thermal Issues

(F1), Overclocking (F2), the construction of New Computer Systems (F3), Operating System (OS) issues (F5), and Memory Speed issues (F6). Overlooking the potential differences here across these themes, while focusing only on the seemingly obvious AMD versus Intel (F4) theme, could thus represent a missed opportunity to target a significant group of lurkers with a defined interest. Second, while classifying threads based on their predominant theme is necessary to plot them, threads can still load on more than one theme. As a result, the impact of certain themes may be understated or overstated in a simple plot if these correlations are not accounted for. For example, the distinct threads associated with Memory Speed issues (F6) may be present because they involve discussions of Memory Speed differences associated with Intel versus AMD products. In which case, the plot may be overstating the level of interest in Memory Speed issues (F6). Finally, replying to a thread requires viewing it, and so an accurate assessment of the relationship between themes and views necessitates controlling for the number of replies. Taken together, it is clear that while examining the scatterplot may hint at which themes are more important to lurkers, this seemingly intuitive approach might lead to spurious conclusions.

To ensure that the data are interpreted correctly, it is necessary to conduct additional tests which verify that: 1) there are true differences across the themes, 2) the influence of multiple themes per thread on *Views* is being accounted for, and 3) these underlying themes are attracting lurkers independently of the replies. To address all of these issues, we conducted a mediation analysis using structural equation modeling based on the conceptual model of the TDA process shown in Fig. 1. The full model tested is shown as Fig. 3. We tested both a fully mediated model, in which *Views* could only be accounted for by *Replies*, as well as a partially mediated model, in which themes could directly affect *Views*, while controlling for the impact of *Replies*. The resulting models containing the six themes are presented in Fig. 3. Each theme is a latent variable with the influence scores of the words serving as the measures. The latent variables are scaled by setting the variance of each to 1. The models were estimated with EQS 6.1 using robust maximum likelihood estimation.

Discovery oriented TDA focuses on revealing the most prominent themes which are of interest to managers. Consistent with this focus on prominent themes, the global fit statistics for both models—the fully (e.g., themes only have indirect effects on *Views* through *Replies*) and partially mediated (the themes have both direct and indirect effects on *Views*)—show good fit when parsimony is accounted for. For the full model, while the non-parsimony adjusted fit statistics indicate poor fit (Satorra–Bentler $\chi^2[1367] = 4830.3$, $p < .001$, CFI = .59), the parsimony adjusted RMSEA statistic indicates very good fit (RMSEA = .021, 90% confidence interval of .020 to .022). The partially mediated model shows the same pattern with the Satorra–Bentler χ^2 and CFI statistics indicating poor fit and RMSEA indicating very good fit (Satorra–Bentler $\chi^2[1361] = 4826.7$, $p < .001$, CFI = .59, RMSEA = .021, 90% confidence interval of .020 to .022). The RMSEA statistic suggests that both models achieve a remarkable and extremely high degree of fit with the data, given their parsimony and emphasis on the most prominent themes (Hu and Bentler 1999).

To identify discrepancies associated with these themes, the standardized path coefficients based on the robust estimation method were examined. The standardized coefficients for both the fully mediated model and the partially mediated model are reported in Table 2. In both models, the standardized coefficient between *Replies* and *Views* is significant, with a value of .78 in the fully mediated model and .81 in the partially mediated model. In the fully mediated model, only the paths from Overclocking (F2) and Intel vs. AMD (F4) to *Replies* are significant with standardized values of .04 and .29 respectively. In the partially mediated models, these paths are also significant and have the same standardized values. However, only one direct path between one of our six themes and the *Views* variable is significant, Intel vs. AMD (F4), with a value of .12. This comparison of standardized path coefficients is an approach recommended by Baron and Kenny (1986) and supports the assertion that the Intel vs. AMD theme (F4) is driving *Views* independent of *Replies*.

Consistent with our reasoning, we found that both F2 and F4 factors have significant direct effects on *Views* (V2) ($p < .05$). When *Replies* (V1) is added into the model as a mediator, the direct effects of F2 and F4 on *Replies* were significant ($p < .05$). There was also a significant effect of the mediator, *Replies* (V1), on the dependent variable, *Views* (V2) ($p < .05$). When the direct path is controlled for, the path between Overclocking (F2) and Views becomes nonsignificant. This indicates that the Overclocking (F2) theme does not garner additional Views; instead, the effect of this theme is fully mediated by Replies. However, the direct path between Intel vs. AMD (F4) and *Views* remains significant, $p < .05$. This suggests that discussions of AMD versus Intel products generate more views above and beyond what is accounted for by the WOM participants, consistent with the discrepancy revealed in Fig. 1. It is also noteworthy that New Computer Systems (F3) and Memory Speed issues (F6) have no significant effects on *Views* (V2), but when *Replies* (V1) is added into the model as a mediator, these themes have significant effects on *Replies*. This suggests that these two themes are likely to be more of interest to active participants than to the general community. Additionally, Thermal Issues and Operating System Issues themes have no significant effects on either *Replies* or *Views*, suggesting that these themes are not driving activity on the forum.

Discussion

The results of the scatterplot in Fig. 2 and the mediation analysis in Fig. 3 demonstrate the efficacy of TDA in discriminating between those themes that generate a discrepant number of views by lurkers from those that do not. For example, in the case of the Overclocking (F2) theme, the relationship is fully mediated through WOM activity (V2). This indicates that the views of webpages and threads associated with this topic are being driven by interest from the active participants rather than lurkers. However, for the AMD/Intel comparison theme, the relationship is a partially mediated one. In this case, while the active participants view these pages, there are a significant number of views being

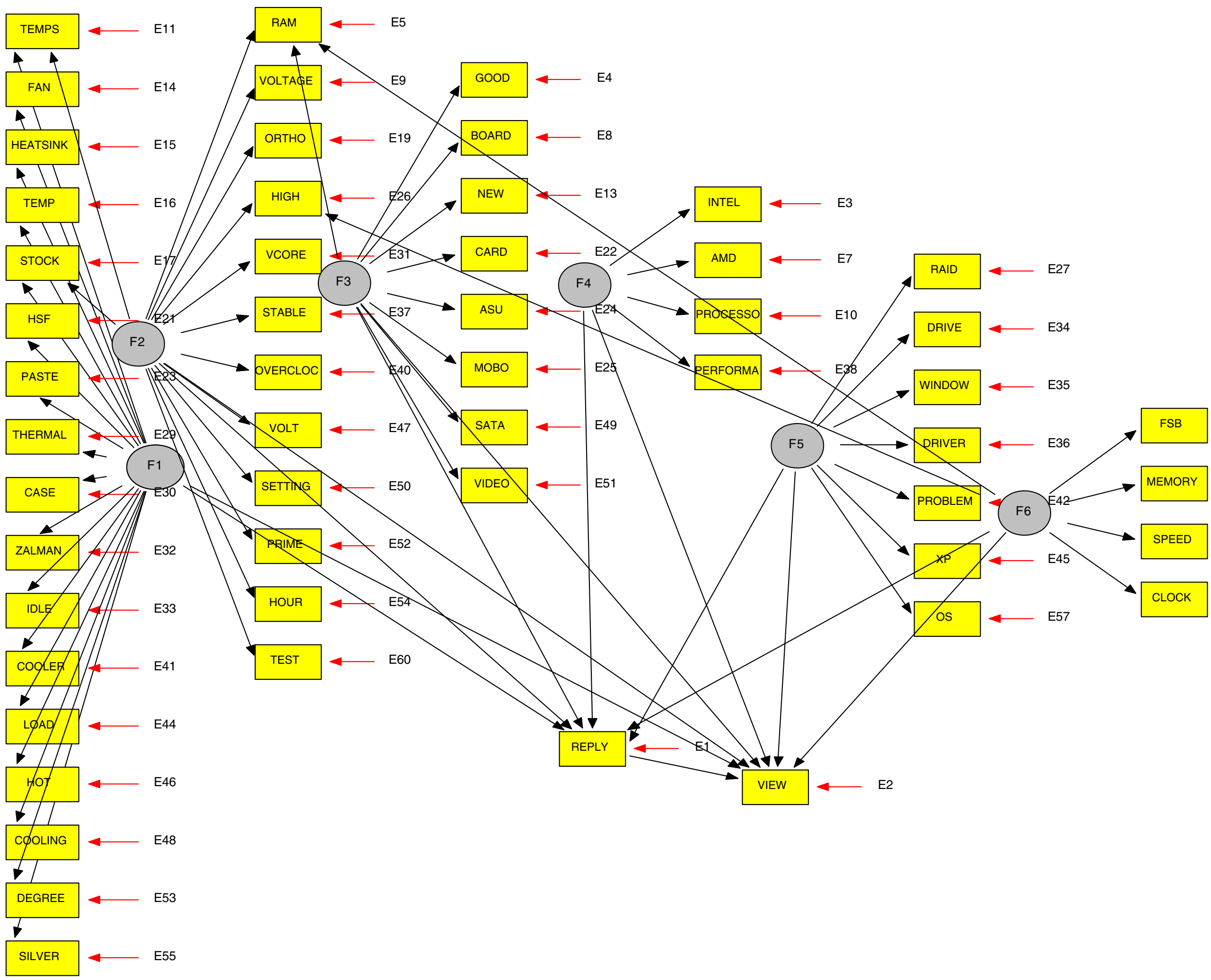

Fig. 3. Partially mediated model.

generated by lurking non-participants that cannot be accounted for based on activity by the participants. Thus, non-participants to this forum are particularly interested in WOM messages that compare the merits of Intel products to its chief rival's products. From a managerial perspective, the non-significant effects of the other four themes are also noteworthy. While these non-significant paths do not mean that these themes are unimportant, they do reveal that these themes are not of disproportionate interest to lurkers. Thus, managers can use the level of interest among the participants as a proxy for the concerns of non-participating consumers, without fear of under- or over-representing their importance.

Based on this insight, the discovery oriented TDA suggests that Intel should focus on comparison based ads that emphasize the advantages of its products over AMD. While this would be "preaching to the choir" for the active participants, this is the information the vast majority of lurkers are seeking. But the results are even more interesting for managers employed by Intel's chief competitor: AMD. Based on the active participants, it appears to be a waste of resources for AMD to purchase advertising on an Intel enthusiast forum. However, the discovery oriented TDA reveals that non-participants are actually visiting this forum to seek competitive information on AMD products versus Intel products. Thus, AMD would benefit from comparison based advertising on *Intel's brand community* since, while the participants may not be receptive, there is a large number of lurkers looking for this exact

Table 2
Standardized path coefficients.

	Fully mediated model		Partially mediated model	
	Themes	Coefficient	Themes	Coefficient
Replies (V1)	F1 thermal issues	.01	F1 thermal issues	.01
	F2 overclocking	.04 *	F2 overclocking	.04 *
	F3 new computers	.03	F3 new computers	.03
	F4 AMD vs. Intel	.29 *	F4 AMD vs. Intel	.29 *
	F5 O/S issues	.00	F5 O/S issues	.00
	F6 memory issues	.02	F6 memory issues	.02
Views (V2)	V1 reply	.78 *	V1 reply	.81 *
			F1 thermal issues	.01
			F2 overclocking	.01
			F3 new computers	.02
			F4 AMD vs. Intel	.12 *
			F5 O/S issues	.05
			F6 memory issues	.01

* $p < .05$.

information. This demonstrates the market-sensing capacity of TDA to reveal opportunities that may be missed if only active WOM participants are considered.

Using TDA to Detect Discrepancies for Pre-existing Topics

The TDA technique may further be applied to verify if there are discrepancies between participants and non-participants on a particular topic. The presence of a discrepancy would indicate a difference between the two, with the sign indicating the nature of a discrepancy. Specifically, a positive sign indicates a higher level of interest among lurkers compared to participants, while a negative sign indicates a lower level of interest. This analysis would typically be conducted for preselected terms of importance to the manager to assess the degree to which the WOM activity around a theme is representative of the larger population of participants and interested non-participants. To test this, we draw on the content analysis and SEM conducted above.

We first tested whether the WOM participants are more interested in OS issues (F5) than non-participants. The non-dictionary based content analysis identified the threads which discuss the topic of interest, OS issues. Examining Fig. 2, there is a notable lack of threads associated with OS issues (F5), among the threads with the highest number of views or replies. This suggests that this theme is not suffering from non-response bias. To determine whether this absence may be due to a failure to control for presence of other themes, we turn to the SEM analysis. In Table 2, the factor associated with this theme is labeled F5, and the standardized coefficients for the fully mediated model and the partially mediated model are reported. By examining the paths, we sought to understand whether the topic of OS issues had varying levels of interest among participants and lurkers. A significant path would indicate a difference between the two groups, and thus a non-response bias. In the fully mediated model, the path between F5 (OS Issues) and V1 (*Replies*) is not significant at the $p < .05$ level. Similarly, in the partially mediated model, the path between F5 and V1 is not significant. More importantly, the path between F5 and V2 (*Views*) is not significant at the $p < .05$ level.

Next, we applied a confirmatory approach to our earlier findings from the exploratory TDA that tested whether there was a difference between the interest level of WOM participants and lurkers when it came to competitive comparisons (F4). The results in Fig. 2 and Table 2 suggest that there is in fact a non-response bias in WOM involving competitive comparisons. In Fig. 2, threads associated with Intel vs. AMD (F4) comparisons garner discrepant levels of *Views*. Supporting this, the path between F4 (Intel vs. AMD) and V2 (Views) in Table 2 is significant in the partially mediated model. Thus, this theme generates a discrepant number of views from lurkers. Furthermore, the sign is positive, indicating that lurkers have a higher level of interest in the theme than do participants.

Discussion

The results of our model show that OS issues, one of the top six themes identified in the initial non-dictionary based content analysis, do not lead to a discrepant number of views. This indicates that non-participants are no more prone to view such content than are WOM participants. As a result, concerns over non-response bias related to this topic are not supported.

In contrast, for the other identified theme of Intel/AMD comparisons, the researcher would conclude that there is evidence of non-response bias, as there is a significant effect in our model. As a result, caution would have to be used when drawing any conclusions or generalizations that are not limited to the active WOM participants.

Taken together, these two confirmatory TDAs illustrate the need for researchers to test for non-response bias on the specific subject matter they are examining. Simply having a large number of lurkers does not mean that the researcher will encounter a significant response bias. When it came to OS issues, there was in fact no evidence of bias. However, when it came to competitive comparisons, we clearly saw a difference between the active participants and the lurkers.

In addition, the results are valuable for managers seeking to identify contexts in which to place specific ads or marketing messages. The results show that consumers turn to this context when attempting to decide between the two competing brands, Intel and AMD. Such contexts would be prime locations for placing advertisements that specifically address the competitive advantages of the firm's products. Additional confirmatory TDAs could be performed on other forums to identify other environments in which competitive comparison ads may be particularly effective.

Sensitivity of TDA to Controversy

A possible concern is how the presence of controversy in the threads could influence results. For example, it could be argued that lurkers may simply be attracted to those threads or themes that contained heated discussions over controversial topics, like children attracted to a schoolyard fight. If controversy is a pervasive topic, a thematic analysis would reveal a separate theme containing the emotional laden terms associated with it. For example, the presence of theme comprised of colorful euphemisms or slanderous name calling would suggest that controversy was playing an important role. Furthermore, a scatterplot followed by a mediation analysis would reveal whether controversy, in and of itself, was driving the interest of lurkers.

However, what if controversy played a more subtle role? It is possible that certain themes may be more subtly controversial, with some emotional laden discussions, but not sufficient for controversy to be identified as a separate theme. To assess the impact of this on TDA's results, we conducted a follow-up analysis.

First, we used the Linguistic Inquiry and Word Count (LIWC) dictionary to identify the number of positive and negative emotionally laden words each thread contained. LIWC has been widely used in the psychology literature to study the affective content of writings (Burke and Dollinger 2005; Mehl and Pennebaker 2003). More importantly, the LIWC dictionary is unique in that its categories have been validated by judges rating hundreds of files of text and has been updated and revalidated periodically for over a decade (Pennebaker et al. 2007). Based on this prior research, the LIWC dictionary includes categories comprised of terms that are associated with positive and negative affect. Next, we conducted an ordinary least squares (OLS) regression analysis with the number of views as the dependent variable and including the number of positive emotional terms, the number of negative emotional terms, the number of replies, and the factor loadings for the six theme factors as predictors. If TDA is detecting subtle differences in controversy, instead of lurker interest in the identified themes, the impact of the AMD vs. Intel theme should be rendered insignificant.

The resulting model showed overall good fit (R^2 = .621, Adjusted R^2 = .620). Consistent with the SEM results, a higher number of replies leads to more views (β = 76.63, p < .01). Furthermore, even when controlling for the impact of emotionally laden discussions, the Intel vs. AMD theme is the only one statistically significant at p = .10, matching the results from Fig. 2 and Table 2. Interestingly, the presence of negatively laden terms does not attract views, but rather decreases them (β = −30.75, p < .01). In contrast, positive terms lead to a slight increase in views (β = 4.07, p = .082).

These results suggest that TDA is robust to the presence of controversy. Furthermore, this robustness attests to the ability of this technique to filter out the noise and bandwagon effects that would be generated by bickering among participants who espouse one view over another. Indeed, to the degree to which a topic involves heated arguments along with the concomitant increase in replies, lurkers are actually repelled, leading to a lower discrepancy rather than a higher one. This robustness makes intuitive sense, as lurkers are seeking information on topics of interest. Participants arguing over controversial topics do not provide clear information that lurkers can act upon. As a result, they avoid such threads in favor of those providing clear, non-controversial information. Thus, in these types of forums, the discrepancies identified by TDA are not simply the product of lurkers seeking out subtly controversial themes.

General Discussion

Thematic discrepancy analysis involves three stages: (1) collecting and analyzing counts of *Views* (WOM impact) and of *Replies* (WOM activity); (2) collecting and analyzing the content of the population of messages to identify themes using non-dictionary or dictionary based approaches; and (3) identifying themes which show discrepancies between WOM activity (*Replies*) and WOM impact (*Views*). Thus, TDA reveals topics of greater interest to lurkers. In doing so, it provides vital insights for both marketing managers and researchers.

To date, there has been little research on methods that can discover and track the topics of interest to lurkers. The lack of such methods has led the WOM literature to focus on analyzing participants while largely ignoring the vast majority of people who read WOM without participating. TDA fills this gap by providing managers with market-sensing information on issues of importance to lurkers that might otherwise be overlooked when just reading WOM generated by participants. In addition, TDA provides insights into marketing messages that may appeal more to lurkers as well as indicating on which websites managers should place advertisements containing those messages. By applying TDA to large scale online WOM data, we demonstrate the ability of TDA to identify themes and content areas generating interest among lurkers, as well as those content areas generating differing levels of interest between lurkers and active participants. Given that there are no other publicly available methods for generating these types of insights about non-participants, this represents a significant contribution to both academic research and managerial practice.

Furthermore, non-response bias represents a critical problem for researchers relying on online WOM data. While this data source provides rich opportunities, researchers have lacked a means to determine whether the interests of WOM participants reflect those of the vast majority of silent lurkers reading it. As a result, generalizing beyond the small number of participants has been problematic. TDA contributes to this literature by providing researchers with a flexible method to evaluate non-response bias for a given topic under study. We demonstrate how TDA can be used to test a pre-determined topic of interest and determine whether it is attracting a discrepant level of interest from lurkers.

A major strength of TDA is its analytic flexibility. TDA is independent of the content analysis package used. Thus, managers and researchers can conduct TDA leveraging the content analysis packages they are familiar with. The only limitation is that the type of content analysis should be appropriate for the purpose of the TDA being conducted. For confirmatory TDA, either dictionary or non-dictionary based approaches are viable. However, for discovery oriented TDA, we recommend non-dictionary content analysis approaches in order to produce the best insights. Finally, TDA is independent of the statistical method used to detect discrepancies. In this paper, structural equation modeling was combined with Baron and Kenny's (1986) approach for testing mediation. However, researchers could use a range of statistical approaches to test the relationships between the themes, views, and replies. This analytical flexibility is a major strength—allowing researchers and managers to choose the approaches that they are most familiar with or that are most aligned with their research objectives.

Managerial Implications

For practitioners, TDA offers a straightforward method to increase a firm's market-sensing capabilities. Given the growing usage of Internet advertising and banner ads, TDA offers a promising means to target advertising within social environments such as forums and message boards. Online forums provide an interesting and cost-effective means to target advertising to individuals who are actively seeking specific types of information. Using TDA, managers can select and customize marketing messages that appeal to the unique interests of non-participants, who represent a larger number of consumers. In doing so, TDA helps managers to discover surprising or counterintuitive opportunities, such as the opportunity to reach nonbrand loyalists by posting banner ads in a competitor's brand community forum.

Furthermore, managers can use insights generated by TDA to proactively identify issues non-participants are seeking information about and provide assistance (through marketing communications and the like) to them. From a strategic viewpoint, this could be especially important when firms are attempting to allay uncertainties that consumers feel about new products (Castaño et al. 2008), as firms could determine which aspects or attributes of a new product are generating the most lurker interest (or concern) and tailor marketing messages to specifically address those needs or issues. For example, managers can use confirmatory TDA analysis to determine whether particular product features are attracting disproportionate interest. They could then focus their messaging on explaining the benefits of these features and how to properly take advantage of them.

Our approach used forum data amassed over a three-year period in one specific forum, which matches the product cycle in this particular industry. Thus, the results demonstrate that TDA is capable of identifying discrepancies among themes that persist over the full product lifecycle of products, rather than just reflecting short term aberrations due to individual marketing actions. However, if managers desired more immediate data, for example, to gauge the effectiveness of an ad campaign or to see the impact of a commercial misstep, they could still apply the TDA approach but across several different sites to ensure the total volume of traffic was sufficient.

TDA also provides insights managers can use to dynamically update their marketing messages to capture changes in consumer interests over time. Castaño et al. (2008) found that consumer concerns regarding the attributes of upcoming new products change over time. TDA provides managers with a means to detect changes in interests among non-participants and, in response, alter the content of promotional messages to emphasize those attributes both expert and novice consumers are most interested in.

Finally, TDA provides managers with a means to detect product complaints that reflect an emerging problem by revealing those concerns that attract a disproportionately high level of views. This allows managers to sift through the massive volume of online WOM and focus only on these complaints that are shared by, or of particular interest to, the wider population of lurkers. As our additional analyses demonstrate, it is important for managers who administer these forums to monitor for aberrant or divisive behavior, as this undermines interest in threads among lurkers and the general body of participants alike.

In summary, the TDA approach provides a powerful tool for managers seeking an easily implementable and relatively inexpensive means of monitoring the themes that are important to consumers before problems emerge or opportunities are lost to competitors. Because the TDA approach relies on the activities of both participants and lurkers alike, this approach provides a degree of insight which is not possible (or financially feasible) using traditional marketing research approaches. TDA also offers a straightforward and flexible method for academic researchers employing online WOM data in their research to test for non-response bias. Given the lack of other approaches to address non-response bias due to lurkers, as well as the increasing use of online WOM data, the TDA approach provides a significant contribution to research methodology.

Future Research and Limitations

Thematic discrepancy analysis opens new avenues for future research. In this paper, we focused on using TDA to examine discrepancies in interest between WOM participants and non-participants toward different topics. In doing so, we found that lurkers showed a particular interest in comparisons of Intel versus AMD products. However, TDA could be extended to also look at discrepancies in WOM valence. In this case, TDA could be employed to determine whether negative or positive WOM about the competing products is attracting more interest from non-participants—and to determine the nature of this interest.

Another interesting area for future research would be to use TDA to examine changes in themes of interest over time and to study the impact that these thematic changes have on sales. While researchers have examined the impact of WOM on sales in the initial stages of product launches (e.g., Luo 2009), less research has been conducted on how changing interests in the WOM topics that are read may impact sales over time. Using TDA, researchers can track changing interests in features, attributes, and concerns not just among WOM participants but among non-participants. This would provide new insights into how the WOM that is *viewed* influences sales, above and beyond what WOM messages are *posted*.

Although TDA provides insights into what lurkers are interested in and how these interests differ from WOM participants, it does not reveal *who* these lurkers are or *why* they differ. When discrepancies or non-response bias problems are detected, follow-up research is required to determine why differences exist and whether or not these differences are theoretically or managerially relevant. At this point, managers must apply their knowledge of their target consumers and their product markets to determine if these discrepancies are centered around themes that warrant additional managerial scrutiny.

References

Algesheimer, René and Utpal M. Dholakia (2006), "Do Customer Communities Pay Off?," *Harvard Business Review*, 84, November, 26–30.

Alterman, Richard and Lawrence A. Bookman (1990), "Some Computational Experiments in Summarization," *Discourse Processes*, 13, 2, 143–74.

Aribarg, Anocha and Natasha Zhang Foutz (2009), "Category-Based Screening in Choice of Complementary Products," *Journal of Marketing Research*, 46, 4, 518–30.

Armstrong, J. Scott and Terry S. Overton (1977), "Estimating Nonresponse Bias in Mail Surveys," *Journal of Marketing Research*, 16, August, 396–402.

Baron, Reuben M. and David A. Kenny (1986), "The Moderator–mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations," *Journal of Personality and Social Psychology*, 51, 6, 1173–82.

Blair, Edward and George M. Zinkhan (2006), "Nonresponse and Generalizability in Academic Research," *Journal of the Academy of Marketing Science*, 34, 1, 4–7.

Burke, Philip A. and Stephen J. Dollinger (2005), "'A Picture's Worth a Thousand Words': Language Use in the Autophotographic Essay," *Personality and Social Psychology Bulletin*, 31, 4, 536–48.

Carley, Kathleen M. (1997), "Network Text Analysis: The Network Position of Concepts," In: Roberts Carl W., editor. *Text Analysis for the Social Sciences: Methods for Drawing Statistical Inferences from Texts and Transcripts.* Mahway, NJ: Lawrence Erlbaum Associates, 79–100.

Castaño, Raquel, Mita Sujan, Manish Kacker, and Harish Sujan (2008), "Managing Consumer Uncertainty in the Adoption of New Products: Temporal Distance and Mental Simulation," *Journal of Marketing Research*, 45, June, 320–36.

Cook, Thomas D. and Donald T. Campbell (1979), *Quasi-experimentation: Design and Analysis Issues for Field Settings*. Boston, MA: Houghton Mifflin Company.

Corman, Steven R., Timothy Kuhn, Robert D. McPhee, and Kevin J. Dooley (2002), "Studying Complex Discursive Systems: Centering Resonance Analysis of Organizational Communication," *Human Communication Research*, 28, 2, 157–206.

Dillman, Don A. (2000), *Mail and Internet Surveys: The Tailored Design Method*. 2nd ed. New York: John Wiley and Sons.

Dwyer, Paul (2007), "Measuring the Value of Electronic Word of Mouth and its Impact in Customer Communities," *Journal of Interactive Marketing*, 21, 2, 63–79.

Godes, David and Dina Mayzlin (2004), "Using Online Conversations to Study Word-of-Mouth Communication," *Marketing Science*, 23, 4, 545–60.

Google (2012), "About Analytics Alerts (Intelligence Events)," (Retrieved Feb. 27, 2013, from http://support.google.com/analytics/bin/answer.py?hl=en&answer=13204912012.)

Grégoire, Yany, Daniel Laufer, and Thomas T. Tripp (2010), "A Comprehensive Model of Customer Direct and Indirect Revenge: Understanding the Effects of Perceived Greed and Customer Power," *Journal of the Academy of Marketing Science*, 38, December, 738–58.

Hennig-Thurau, Thorsten, Kevin P. Gwinner, Giafranco Walsh, and Dwayne D. Gremler (2004), "Electronic Word-of-mouth via Consumer-opinion Platforms: What Motivates Consumers to Articulate Themselves on the Internet," *Journal of Interactive Marketing*, 18, 1, 38–52.

Hickman, Thomas and James Ward (2007), "The Dark Side of Brand Community: Inter-group Stereotyping, Trash Talk, and Schadenfreude," *Advances in Consumer Research*, 34, 314–9.

Hu, Li-Tze and Peter M. Bentler (1999), "Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus New Alternatives," *Structural Equation Modeling*, 6, 1, 1–55.

Humphrey, Suzanne M. (1999), "Automatic Indexing of Documents from Journal Descriptors: A Preliminary Investigation," *Journal of the American Society for Information Science*, 50, 661–74.

Keller, Kevin L. (2003), *Strategic Brand Management: Building, Measuring, and Managing Brand Equity*. Upper Saddle River, NJ: Pearson Education.

Kozinets, Robert V. (2002), "The Field Behind the Screen: Using Netnography for Marketing Research in Online Communities," *Journal of Marketing Research*, 39, February, 61–72.

———, Kristine de Valck, Andrea C. Wojnicki, and Sarah J.S. Winer (2010), "Networked Narratives: Understanding Word-of-Mouth Marketing in Online Communities," *Journal of Marketing*, 74, March, 71–89.

Liu, Yong (2006), "Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue," *Journal of Marketing*, 70, July, 74–89.

Luo, Xueming (2009), "Quantifying the Long-term Impact of Negative Word-of-mouth on Cash Flows and Stock Prices," *Marketing Science*, 28, 1, 148–65.

Mathwick, Charla (2002), "Understanding the Online Consumer: A Typology of Online Relational Norms and Behavior," *Journal of Interactive Marketing*, 16, 1, 40–55.

Mehl, Matthias R. and James W. Pennebaker (2003), "The Sounds of Social Life: A Psychometric Analysis of Students' Daily Social Environments and Natural Conversations," *Journal of Personality and Social Psychology*, 84, 4, 857–70.

Morrissey, Brian (2010), "Display Ads Try for a Banner Year," *Adweek*, 51, 6, February 8, 6.

Muñiz Jr., Albert M. and Thomas C. O'Guinn (2001), "Brand Community," *Journal of Consumer Research*, 27, 4, 412–32.

——— and Hope Jensen Schau (2005), "Religiosity in the Abandoned Apple Newton Brand Community," *Journal of Consumer Research*, 31, 4, 737–47.

Nielsen, Jakob (2006), "Participation Inequality: Encouraging More Users to Contribute," (Retrieved Feb. 27, 2013, from http://www.nngroup.com/articles/participation-inequality2006.)

Pennebaker, James W., Matthias R. Mehl, and Kate G. Niederhoffer (2003), "Psychological Aspects of Natural Language Use: Our Words, Our Selves," *Annual Review of Psychology*, 54, 547–77.

———, Cindy K. Chung, Molly Ireland, Amy Gonzales, and Roger J. Booth (2007), *The Development and Psychometric Properties of LIWC2007*. Austin, TX: LIWC.net.

Rogers, Everett M. (2003), *Diffusion of Innovations*. New York: Free Press.

Sheng, Ellen (2007), "Corporate Connections: Companies Find Social Networks Can Get People Talking—About Their Products," *The Wall Street Journal*, R8 ((January 29)).

Thompson, Scott A. and Rajiv K. Sinha (2008), "Brand Communities and New Product Adoption: The Influence and Limits of Oppositional Loyalty," *Journal of Marketing*, 72, October, 65–80.

Tran, K. (2000), "Group Think: Community Connect Has Created Ethnic Associations on the Net; Can It Also Make Money?," *The Wall Street Journal*, R53 ((October 23)).

W3Techs (2011), "Usage Statistics and Market Share of vBulletin for Websites," (Retrieved Feb. 27, 2013, from http://w3techs.com/technologies/details/cm-vbulletin/all/all2011.)