Marketing Science Institute Working Paper Series 2011
Report No. 11-116

No Comment?! The Drivers of Reactions to Online Posts in Professional Groups

Robert P. Rooderkerk and Koen H. Pauwels

Report Summary

Social media has moved beyond personal friendships to professional interactions in high-knowledge industries. In such settings, online discussion forums are sponsored by firms aiming to position themselves as thought leaders, to gain more insight in their customer base, and to generate sales leads. However, while firms can seed discussion by posts, they depend on the forum members to continue the discussion in the form of reactions to these posts.

Here, authors Rooderkerk and Pauwels investigate what features and characteristics determine the number of comments that a post receives on an online discussion forum. The study is the first in-depth statistical analysis of the behavior of members of the focal discussion group.

The empirical setting involves a global manufacturer connecting with health care professionals through a LinkedIn discussion forum. The authors project that content characteristics, post characteristics, author characteristics, and timing characteristics will jointly determine the number of comments a post receives.

They tested their conceptual framework by analyzing a collection of 316 threads; i.e., a post (opening article) and comments (reactions). The independent variables that could not be observed directly were judged by two human raters with sufficient expertise in the corresponding domain. Using count data models, the authors established the effects of the different types of characteristics on the number of comments.

Findings

The authors found that more controversial topics and topics with practical utility result in more comments. The self-centeredness of the post and the ambiguity of the topic did not affect the amount of comments significantly.

With respect to post characteristics, the length of the post negatively affected the amount of comments. Sentence length did not have a significant effect. Readability had a significant positive effect on the amount of discussion following a post. The emotionality of a post and the use of jargon had no substantial effect. Explicitly phrasing a post as a question and encouraging members to respond both increased the number of comments. Posting on a weekend significantly reduced the number of following comments, as did including a hyperlink in the post.

Socio-economic status was the only author characteristic that significantly increased the number of comments a post received; there was no evidence for the effect of the author's gender. The number of author connections (i.e., author popularity) did not have any effect on the number of comments.

These results provide valuable insights for firms on how to build and maintain an online forum through ongoing discussions.

Robert Rooderkerk is Assistant Professor and CentER Fellow at the Tilburg School of Economics and Management, Tilburg University, Tilburg, The Netherlands. Koen Pauwels is Professor of Marketing at Özyeğin University, Istanbul, Turkey.

Acknowledgments
The authors thank Patrick Filius of Philips Healthcare for providing access to the data and his continued involvement throughout the study. They also thank Jori van de Spijker for assistance with the data collection and coding. In addition, they thank Berk Ataman, Dominique Hanssens, and Marnik Dekimpe for helpful comments and suggestions. The authors gratefully acknowledge the Marketing Science Institute for research support (MSI grant #4-1721).

"Philips is building true engagement with an important customer base by creating meaningful communities on LinkedIn, communities which allow members to collaborate and share insights delivering real value. They've truly grasped the potential of what it means to be driving professional conversations in the right context as a way to increase brand equity and, as such, should be recognized as global thought leaders in this area."
Jeff Weiner – CEO LinkedIn

"By facilitating the dialogue around clinical content, we are able to engage with our target audience, and to enhance our reputation and increase our brand preference."
Patrick Filius – Global Online Marketing Manager, Philips Healthcare
Kors van Wyngaarden – Global Director Online Marketing, Philips Healthcare

Introduction

In recent years the Internet has evolved into a dynamic network where people can easily and constantly connect with each other (Cheung et al., 2008; Stephen and Lehmann, 2009a). Social media websites allow consumers from around the world to interact and inform each other with regard to products and services (Stephen and Toubia, 2010). Increasingly, B2B firms embrace social media as a way to connect with their professional clients. Often their initiatives take the shape of establishing online communities. Firm goals include the (i) positioning as thought leader in knowledge-intensive industries, (ii) gaining insights used for product innovation, (iii) developing meaningful relationships with the customer base and (iv) increasing brand preference resulting in sales leads (LinkedIn 2010a-b-c, 2011a-b). Currently, B2B firms selling products spend on average 4.2% of their marketing budget on social media (The CMO Survey 2011). They plan on increasing this to 7.0% in the next two years and 13.4% in the next five years.

To build and maintain attractive forums it is crucial for firms to stimulate discussion appealing to the forum members. The presence of such discussions helps to position the firm as a thought leader, while absence of online participation may be detrimental to such positioning. Wiertz and De Ruyter (2007) argue that the success of firm-hosted commercial online communities entirely depends on the willingness of the users of the platform to spend time and effort responding to each other. Online discussion forums share this need for member investment with other social media, such as microblogging (e.g. Twitter) and social networks (e.g. Facebook) (Hoffman and Fodor 2010). However, under-contribution is a problem for many online communities (Beenen et al. 2004), as encouraging participation has proved to be one of the greatest challenges for any online community provider (Bishop 2007). Discussions are started by a member posing a specific question or statement. We refer to these discussion starters as posts. It is not clear what posts lead members to engage in discussion. That is, insights are needed into which post characteristics encourage members to comment. Specifically, in this study we set out to answer the following *research question*: *which content, post, author, and timing characteristics increase the number of comments a post receives?*

Interesting to researchers, our findings also lead to actionable recommendations for firms running a forum. First, firms can highlight the content, post and author most likely to get comments. Second, many firms hire communication agencies to start discussions on their online discussion forum. Optimizing the design and content of the topics that are inserted in the forum should lead to more discussion in the form of comments. Finally, the newfound knowledge might also be used in future social media activities (i.e. corporate blogs). In sum, our results can help firms to grow their online discussion groups.

Recent research has shed light on the motivations for consumers to engage in social media (Hoffman and Fodor 2010), and the consequences of social media use by consumers (Chen and Xie 2008; Chevalier and Mayzlin 2006; Dellarocas 2003; Godes and Mayzlin 2004; Godes and Mayzlin 2009). Other studies analyzed the value of online word-of-mouth (Libai et al. 2009; Trusov et al. 2009, 2010). The current study distinguishes itself from previous research, by investigating *how much social media participation is generated by specific posts.* We contribute to the academic literature by investigating which post characteristics encourage professionals in a

B2B industry to engage in online discussions. Despite the recent spur in interest, surprisingly little is known about what drives individuals to engage in specific online conversations (Stephen and Lehmann, 2009b).

The remainder of this paper is organized as follows. In the next section we discuss social media in general and online discussion forums, our empirical application, in particular. Next, we present our conceptual framework that postulates the drivers of member participation in discussions on an online forum. The subsequent section describes the methodology used for testing the conceptual model. Next, we present the estimation results, and conclude with their discussion.

Research Background: Social Media and Online Discussion Forums

In this section we define social media and describe its different types. We then zoom in on online discussion forums and explain how the amount of discussion on these platforms is both critical as well as challenging. We end this section with a discussion of our empirical setting, an online forum for healthcare professionals managed within LinkedIn.

Social media

Online discussion forums fall within a broader class of social media. Kaplan and Haenlein (2010) define social media as "a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0 and that allow the creation and exchange of user-generated content." Enclosed in this definition is the relation of social media to web 2.0 and user-generated content. Web 2.0 refers to the development of revolutionary technologies and functionalities, such as Adobe Flash and RSS, which support applications that enable users to continuously modify the content on a website. User-generated content predates web 2.0, as simple discussion forums and newsgroups existed before these technologies were developed.

However, due to technology constraints in web 1.0 the possibilities were limited. Distinct features of social media are interactivity, stimulating conversation on the Internet rather than one-way communication, and the lack of geographical boundaries and time constraints (Kaplan and Haenlein 2010).

Classification of social media outlets

Many online channels qualify as social media platforms. First, we provide an overview of the different types. Next, we highlight the differences based on key characteristics.

Collaborations concern websites such as Wikipedia. They allow the creation of a large number of interlinked web pages to compose a rich source of information. *Social role-playing & online gaming* refers to social media initiatives where users assume the role of a self-designed character to interact with others in an online virtual world. Interaction occurs not only through moving around in the virtual world, but participants also use text chatting and messaging to communicate. Popular examples of this category are Second Life and World of Warcraft. *Multi-media upload-site*s are websites where users can upload and share content that is non-written. Examples are YouTube (videos), Flickr (photos) and SlideShare (presentations). *Weblogs* are ‘frequently updated, reverse-chronological ordered entries on a single Web page’ (Aggarwal, Gopal, and Sankaranarayanan, 2010). There are many types of blogs, differing both in content and in writing style. *Micro-blogging* is a more recent form of blogging. It differs from traditional blogging, as it contains shorter messages (e.g. Twitter has a maximum of 140 characters per post) and posts are more frequent (Java, Finin, Song, and Tseng, 2007). Although Twitter is the most prominent example of microblogging, other initiatives have emerged such as Tumblr and Plurk.

Social network sites are web-based services that allow individuals to construct a public or semi-public profile within a bounded system, articulate a list of other users with whom they are connected and view a list of connections of others (Boyd and Ellison 2007). Popular sites include Facebook, MySpace, LinkedIn and Hyves. *Discussion groups* refer to Internet-based forums and computer-mediated social gatherings. Online discussion groups or forums are defined as 'places in which consumers often partake in discussions whose goals include attempt to inform and influence fellow consumers about products and brands' (Kozinets, 2002). Brown, Broderick and Lee (2007) argue that consumption-related online communities are representations of word-of-mouth networks, where individuals with a shared interest regarding a certain product category interact. These online communities offer an increasingly prominent environment for interpersonal exchange, as it allows members to continuously share opinions (Miller, Fabian, and Lin, 2009). Steyer, Garcia-Bardidia and Quester (2006) highlight that online discussion groups have the potential to be great sources for data collection, as the discussions can be recorded in real time and information is available regarding the source and the sequence of the messages.

There are several characteristics that distinguish the different types of social media. Table 1 describes the aforementioned types of social media based on a number of key characteristics (tables and figures follow References throughout). First, social media platforms differ in the *level of self-disclosure*, concerning the extent to which people are supposed to provide personal data that can identify them. For example when adjusting content on Wikipedia, one remains completely anonymous, whereas social networking sites specifically require users to disclose personal information. Second, social media forms also differ in their use, which may be primarily *entertaining or informing*. Obviously, this can differ across and within platforms. Blogs can vary for example; as some put more emphasis on entertainment, while others aspire to be a source of information. Next, some social media websites require users to have a *personal page* or account, while others are more focused on creating a general page for all users. Multi-media uploading-sites are somewhat in between, as they require users to have an account before they are allowed to upload, but users generally search the main page rather than personal accounts. *Posting frequency* is obviously determined by users of the website, but general inferences can be drawn. Microblogging, for example, requires users to post very frequently in order to stay appealing to readers. Collaborations like Wikipedia on the other hand only need minor adjustments once an article is completed. Finally, *media richness* concerns the ability of a communication medium to reproduce the information sent over it (Daft and Lengel, 1986). For example, a social networking site like Facebook allows users to produce content in multiple forms, such as videos, pictures and written information. Discussion forums on the other hand are much more plain, and do not provide such elaborate media richness.

Online discussion forum on LinkedIn for healthcare professionals

Many companies use the LinkedIn environment to start discussion groups. Examples include British Gas for Business, Cisco, Hewlett-Packard, Philips, and Sage. Using the LinkedIn environment allows firms to benefit from the readily available IT infrastructure while it facilitates access to a large and still expanding global audience. LinkedIn operates the world's largest professional network on the Internet with more than 150 million members in over 200 countries and territories (LinkedIn 2012). In July 2011, LinkedIn surpassed MySpace to become the second largest social network platform in the US after Facebook (Bloomberg 2011).

LinkedIn facilitates members to start groups on specific topics. Online discussion groups on LinkedIn enable the firms to connect with (potential) customers in a relatively inexpensive way. Following Table 1 these groups can be classified as hybrids between a discussion group and a social networking site. In fact, they can be seen as discussion group within a social network domain. More specifically, members have a relatively high level of self-disclosure with a personal page on a platform that is mostly informative with low media richness. Finally the post frequency of members can be seen as relatively low; one of the key threats to the viability of a discussion group is the lack of engagement, or to be more specific, the lack of discussion.

In our empirical application we focus on a LinkedIn discussion group for healthcare professionals. The group, carrying the name "*Innovations* in Health", is established and maintained by Philips. This company is a global manufacturer of, among others, advanced healthcare products such as fMRI machinery. Philips established the group to build thought leadership, engage with the target audience, facilitate peer-to-peer discussions, gain customer insights, and detect product issues early on. The company decided to use the LinkedIn environment for its discussion group as the target audience was widely represented on this social media platform. Leveraging of the LinkedIn expertise and tools allowed Philips to jump start their own social media initiatives. Figure 1 depicts a screenshot of the "*Innovations* in Health" group.

The "*Innovations* in Health" group provides healthcare professionals a platform, hosted by Philips, to connect with their peers. At the time of data collection the group had 16,000 subscribers. Members included doctors (generalists and specialists), technicians and hospital administrators. Nearly 90% of all members originate from the US, UK, the Netherlands and India, with the latter making up less than 5% of the population. When asked what they like best about the group (Philips 2011), members answered:

"Meaningful/interesting discussion about future trends – offshoring of healthcare, applicability of mobile medicine, etc. Unlike other healthcare related groups, they're not just interested in references and job offers."

"Some discussions are really about hot topics and provide interesting contacts."

"Good ideas are generated. Engaged group. Always something to learn about."

The discussion forum is made up of content created by the group members. As concepts such as threads, posts, comments and topics are sometimes used interchangeably it is useful to define them as they will be used in this study. A *post* is an opening article written by someone who wants to start a discussion with other members of the group. Other members reply with their *comments*, which are their written reactions to the opening post. The collection of the opening post and comments together make up a *thread*. A *topic* is defined here as the subject of interest in a thread; e.g breast cancer screening in Figure 1.

A key goal of the global manufacturer sponsoring the forum (Philips) is to be seen as the thought leader in healthcare. It perceives the online discussion group as instrumental in reaching this goal. To this end, the company believes it is crucial to generate a lot of discussion between its members. This can be achieved by members posting a discussion topic with other members responding to it. Whereas there is a steady increase in the number of posts, the number of comments shows room for improvement. In fact, the majority of posts do not evoke a single comment. Consequently, it is an interesting question what factors determine the number of comments a certain post evokes.

Conceptual Development: Drivers of Participation in an Online Discussion Forum for Healthcare Professionals

Most research on what drives people to participate in online discussion forums has focused on the individual motivations to engage in such activity (i.e. Ardichvili et al. 2003; Dholakia et al. 2004; Hennig-Thurau et al. 2004; Wasko and Faraj 2005). Motivations such as concern for others and self-enhancement were found to determine discussion forum participation. More generally, Hoffman and Fodor (2010) discuss connection, creation, consumption and control as drivers of a consumer's use of social media. Also, Wiertz and De Ruyter (2007) argue that certain people have a higher intrinsic propensity to engage in online interaction than others. But these findings do not explain why certain *posts* lead to lengthy discussions, while others languish. The current research wishes to address this issue by exploring the differences between posts in the discussion group and investigating the number of comments they evoke. We will investigate this issue by looking at what is posted (content characteristics), how it is posted (post characteristics), by whom it is posted (author characteristics), and when it is posted (timing characteristic). Next, we discuss the focal characteristics in our study and formulate the corresponding hypotheses. The hypotheses are summarized visually in Figure 2.

Characteristics of the content

An intuitively appealing starting point that might determine the number of comments is the actual content of the post. Berger and Milkman (2012) found that content characteristics of *New York Times* articles influenced the virality of these articles. Similarly, the content of a certain discussion topic might influence the number of comments it receives.

Controversiality. In their interaction analysis model, Gunawardena et al. (1997) propose that any interaction starts with sharing and comparing (e.g. a statement or opinion) and is followed by the exploration and the potential discovery of dissonance. When there is disagreement, people experience a sense of dissonance as they are exposed to conflicting convictions. Cognitive dissonance theory (Festinger, 1957) argues that people are motivated to solve this uncomfortable feeling of contradictory ideas. Controversial topics should therefore evoke more reaction, as people want to address the source of their feeling of dissonance. Therefore, we propose:

H1: More controversial discussion posts will evoke more comments.

Practical utility. Professionals in the healthcare industry joined the LinkedIn group to discuss job-related matters and obtain information that is useful to them in practice. Therefore, it is reasonable to assume that posts that contain practically useful information or have the potential to become practically useful discussions will evoke more comments. These topics should not only be rated higher, but group members might also be more inclined to comment because they can add to the discussion with their practical experience.

H2: Discussion posts containing a more practically useful topic will evoke more comments.

Self-centeredness. A major reason for people to engage in conversations is because they like to interact with others. Indeed, research has found that people engage in online conversations because they enjoy the social aspect of interpersonal connectivity (Dholakia et al., 2004; Hennig-Thurau et al., 2004). Although traditional conversational norms hold that people should aspire to spur conversations that are informative and relevant to others (Grice, 1975) some people actually place more weight on their own interest (Stephen and Berger, 2010). These members use the discussion forum for blatant self-promotion, having no regard for what others are interested in.

As they violate the traditional conversation norms (Grice, 1975) and signal no intent to start actual interpersonal interaction, such self-centered posts should receive fewer comments.

H3: Discussion posts that are more self-centered will evoke fewer comments.

Topic ambiguity. Ambiguous topics induce uncertainty for the post recipients in what exactly the topic starter is talking about. Intolerance for uncertainty or ambiguity is not uncommon in society (Hofstede 2001). Hence, conversational norms hold that people articulate themselves in a direct and straightforward manner, thereby avoiding ambiguity or obscurity (Grice 1975). Consequently, ambiguous posts should receive less attention as they are in violation of the conversational norms.

H4: Discussion posts containing more ambiguity will evoke fewer comments.

Characteristics of the post

Next to the content of the post, external factors might encourage or discourage information sharing (Stephen and Lehmann 2009b).

Post length. Grice's (1975) conversational norms hold that contributions to a conversation should be straightforward and contain only relevant information. Especially on the Internet people filter information and are not used to spending much time on one item. All of the bits and pieces of information on the Internet compete for the visitor's attention, demanding the same limited cognitive resources. Consistent with effort minimization (Johnson and Payne 1985) individuals should prefer posts that demand less cognitive resources. Therefore, long posts should receive fewer comments than short posts.

H5: Longer discussion posts will evoke fewer comments.

Sentence length. Sentences that contain a lot of words typically take longer to read. This reduces the number of comments that the post will evoke. Consistent with the motivation for the previous hypothesis, members should prefer posts with shorter sentences.

H6: Discussion posts that on average contain more words per sentence will evoke fewer comments.

Readability. People are more likely to comment on posts that are easy to read for the same reason they like short posts: they aim to get the maximum with the smallest amount of effort (Johnson and Payne 1985). Providing the topic in a clear format is foremost the task of the topic starter. Therefore, readability might influence the number of comments on a certain topic, with more complex posts receiving fewer comments.

H7: Discussion posts that are easier to read will evoke more comments.

Sentiment. Every bit of information humans are exposed to primes certain associations in memory. That piece of information highlights certain concepts that are stored in the brain (Fiske, 2004). Because people strive to be happy, they tend to look for information that is positive, which would highlight positive associations and induce a positive mood. Contrarily, people try to avoid negative information that could stimulate a negative mood. This is especially the case for people with a promotion focus. In Western countries, overrepresented in our sample, people tend to have more of a promotion rather than a prevention focus (Lee et al. 2000). Hence, members of the online healthcare group can be expected not to participate in an online discussion when it contains a lot of negativity. At the same time, they should be drawn towards positive discussion posts, as these might improve their mood.

H8: Discussion posts that contain relatively much positivity will evoke more comments, compared to neutral posts.

H9: Relative to neutral posts, the more negative a discussion post is, the fewer comments it will evoke.

Use of jargon. Posts differ in the use of jargon (vocabulary specific to the healthcare industry). When relatively much jargon is used only experts on the matter can properly understand and consequently comment on it. We believe this provides an incentive for members to contribute. Members are expected to pick up on the jargon and feel more inclined to comment as they feel they specifically are able to do so.

H10: Discussion posts that contain relatively more jargon will evoke more comments.

Question in the topic title. Often a topic title is all forum users view when scrolling through the posts. A clear question in the topic title might trigger the interest of the reader. Knowing the problem of the topic starter, allows them to judge whether they can be of help or show their expertise, which were found to be core motivations for discussion forum participation (Ardichvili et al., 2003; Hennig-Thurau et al., 2004). As such, an explicit question in the topic title might lure forum members to respond.

H11: Discussion posts containing a specific question will evoke more comments.

Encouragement to reply. Some postings contain an active encouragement. It is expected that encouragement by the author to submit reactions will in fact lead to more comments.

H12: Discussion posts containing encouragement to reply will evoke more comments.

Inclusion of hyperlink(s). Some postings contain one or more links to Internet articles that might be interesting to the readers. However, this requires more effort from the reader. In line with effort minimization (Johnson and Payne 1985) this reduces the likelihood they will reply. In addition, they could simply be sidetracked by the hyperlink and not return to the online discussion forum.

H13: Discussion posts containing a hyperlink will evoke fewer comments.

Characteristics of the author

People are selective transmitters, meaning that they purposely choose to whom they convey information and to whom they do not (Stephen and Lehmann, 2009b). Consequently, characteristics of the author of the post should be taken into consideration as well.

Author popularity. People like to associate with successful others (Cialdini et al., 1976). Also, a person with more connections is more likely to be closely tied to others in the community and should therefore receive more feedback on topics (s)he started. Such forces may play even stronger for professional sites like LinkedIn, which was designed as an online platform to build a network that might provide job-related opportunities. In this context, users may prefer building relations with people who have many connections, as these are more likely to provide attractive opportunities.

H14: Discussion posts by more popular authors will evoke more comments.

Author social-economic status. Stephen and Lehmann (2009b) argue that people decide how to transmit information (e.g. respond with a comment) based on the anticipated social benefits they expect to receive from forming a relation with that person. Consequently, it can be expected that people are more inclined to respond to posts from people they perceive to able to grant them future business opportunities. Again, this tendency is likely more pronounced in a professional network such as LinkedIn.

H15: A higher perceived social-economic status from the discussion post author will evoke more comments.

Timing characteristic

Given the professional nature of the discussion forum, people should spend more time responding to posts during the working week. Discussions started on week days will thus get more attention than discussions started on the weekend. Come Monday, such weekend discussions will have to compete with those that are posted during Monday.

H16: Discussion posts that were placed during the weekend will evoke fewer comments than those placed during a weekday.

Control variables

Next to the explanatory variables we include the author's gender as a control variable. Schler, Koppel, Argamon and Pennebaker (2006) found significant differences in writing style between male and female bloggers. Nowson, Oberlander and Gill (2005) argue that gender differences are projected in the language used in weblogs, with women writing more contextual than men. We do not have a clear expectation about these effects on the number of comments a post generates. Therefore, we include author gender as control variable. In addition, we include a monthly trend to capture long term evolution of commenting.

Methodology

Our methodology needs to account for the characteristics of our data, which consists of a collection of threads, i.e. a post and the comments that follow it. First, the number of comments evoked by a posts is a non-negative integer number (count data). Second, the number of comments across different posts displays overdispersion (i.e. high variability, long tails). We allow for overdispersion by adopting a count data model that assumes the distribution of the underlying data to be Negative-Binomial. In addition, we explicitly test for overdispersion by also estimating a model that assumes the underlying data to follow an equidispersion Poisson distribution. To our knowledge these advanced statistical analyses improve over the current modus operandi in the industry and the focal firm, which is limited to the use of descriptive statistics.

Model formulation

We assume that the number of comments to post i, Y_i, obeys the following Negative-Binomial process:

$$P(Y_i = k) = \frac{\Gamma(\lambda_i/\theta + k)}{\Gamma(\lambda_i/\theta)\Gamma(k+1)}\left(\frac{1}{1+\theta}\right)^{\frac{\lambda_i}{\theta}}\left(\frac{\theta}{1+\theta}\right)^{k},\ \lambda_i > 0,\ \theta > 0, \tag{1}$$

where

Y_i = the number of comments evoked by post i (=1,…,N), with N the number of posts,

$\Gamma(\cdot)$ = the gamma distribution.

The Negative Binomial distribution is a two-parameter distribution. The two parameters are respectively the λ_i and θ. The expected number of comments of post i, E(Y_i) is equal to λ_i. The corresponding variance, Var(Y_i), is equal to $\lambda_i(1+\theta)$. The theta parameter is often referred to as the overdispersion parameter. Larger values for theta represent more overdispersion of the underlying data. When theta approaches zero the negative binomial distribution converges to a Poisson distribution which has equidispersion, meaning that mean and variance are equal (to λ_i). We will also estimate the single-parameter Poisson distribution and compare the models based on fit and complexity.

Next, we relate the lambda parameters to the explanatory variables:

$$\log(\lambda_i) = \beta_0 + \beta_1 \cdot CONTR_i + \beta_2 \cdot PRACT_i + \beta_3 \cdot SELF_i + \beta_4 \cdot AMBIG_i + \\ + \beta_5 \cdot POST\_LENGTH_i + \beta_6 \cdot SENT\_LENGTH_i + \beta_7 \cdot READ_i + \beta_8 \cdot POS_i + \\ + \beta_9 \cdot NEG_i + \beta_{10} \cdot JARGON_i + \beta_{11} \cdot QUESTION_i + \beta_{12} \cdot ENCOUR_i + \beta_{13} \cdot HYPER_i + \\ + \beta_{14} \cdot NUM\_CONNECT_i + \beta_{15} \cdot SES_i + \beta_{16} \cdot WEEKEND_i + \beta_{17} \cdot FEMALE_i + \\ + \beta_{18} \cdot TREND_i. \quad (2)$$

Table 2 provides the definitions of the explanatory variables.

Elasticities

Comparing the effect sizes across variables can best be done by comparing the marginal effect of each variable. To this end we compute the elasticities. For a continuous variable the elasticity is given by (Washington et al. 2003)[1]:

$$\eta_{x_{ij}}^{\lambda_i} = \frac{\partial \lambda_i}{\partial x_{ij}} \cdot \frac{x_{ij}}{\lambda_i} = \beta_j x_{ij}, \quad (3)$$

where x_{ij} is the j^{th} variable in the vector of explanatory variables for post i, and β_j is the corresponding coefficient for the j^{th} variable. In case of a dummy variable we compute the pseudo-elasticity as an approximate elasticity of this variable (Washington et al. 2003):

$$\eta_{x_{ij}}^{\lambda_i} = \frac{\exp(\beta_j) - 1}{\exp(\beta_j)}. \quad (4)$$

Next, we describe the empirical setting to which we apply our model.

Empirical Application

In this section we first describe the data. Next we describe the estimation results.

[1] The metric variables are included in the model in their standardized form. The Appendix shows how the formula in Equation (3) can be adjusted to obtain the elasticity with respect to the unstandardized variable while using the coefficient corresponding to the standardized variable.

Data description.

Sample. LinkedIn was contacted by Philips to obtain the necessary data. During a period of 9 months (October 2009 – June 2010), we observe 316 relevant posts on threads finished before the end of the data period. On average, the number of days until the last comment was inserted was 11.52 days, with the longest thread (i.e., post + corresponding comments) being active for 85 days.

Measurement. For a lot of the variables, such as post length or presence of a hyperlink, measurement is straightforward. However, some of the independent variables cannot be observed directly, they have to be judged by a human rater. To increase objectivity, multiple human raters were asked to judge the same data. Two human raters were found with sufficient expertise in the corresponding domain and in command of the English language. They were unaware of the goals of the study. Moreover, they did not know how many comments the posts generated. Both coders were asked to rate the content characteristics for each post independently of each other. Clear coding instructions were provided. The different subjective dimensions (*controversy, practical utility*, *self-centeredness*, *topic clarity*, *readability, use of jargon*) are rated on a seven-point Likert scale (1 = not at all, 7 = extremely) for each post. The *sentiment* of a post, the *positivity* and *negativity,* is determined using the freely available Linguistic Inquiry and Word Count (LIWC) software. This package performs a content analysis and classifies each word as positive, neutral, or negative. Measures for both *positivity* (% of words that are positive) and *negativity* (% of words that are negative) are part of the LIWC output. *Author popularity* is captured by the number of connections the author had for his LinkedIn profile. This information was not displayed directly next to the post. However, the information is just one click away and checking other members' profiles is common in LinkedIn, especially for people who use the discussion group for networking. The *author's social-economic status* (SES) is coded by human raters on a seven-point Likert scale based on the author's job title. One would expect the perceived SES of a technician at a hospital with little reputation to be lower than that of brain specialist from a reputable hospital. We use a dummy variable to indicate if the post was placed over the weekend. In addition, we use a trend variable at the monthly level to allow for long-term trends in the level of commenting based on the growth of the discussion forum.

Inter-rater reliability. We compute the level of inter-rater reliability at the variable-level using Spearman's rho (correlation) on the ordinal (Likert-type) data. A correlation of .1 to .29 should be considered small, .3 to .49 should be considered medium and .5 to 1.0 should be considered large inter-rater reliability (Cohen, 1988). The correlations for all variables, except for readability, fall within the large inter-rater reliability category. Readability falls into the medium category with a rho of .49. However, it is on the border with the large category. Consequently, we believe the agreement between the judges is sufficiently high. Consequently, we take the averages of their scores to measure the constructs. Table 3 provides the descriptive statistics for the resulting dataset.

Figure 3 represents the distribution of the number of comments across all posts in a histogram. Besides overdispersion, the distribution also reveals that more than 60% of the posts do not evoke a single comment. This highlights the relevance of our research, but also suggests that the data may be zero-inflated; i.e. the fraction of zeros is too high to be compatible with a standard underlying count data model (Winkelmann 2008, p. 173). Theoretically, the process generating the zeros might depend on other factors than the process for strictly positive outcomes. In our study, posts may raise no comments because of some apparent factors (e.g., being extremely lengthy) that may have no or differential impact on a given positive number of comments. To allow for this to occur in our data we also estimate *zero-inflated versions of the*

Negative-Binomial and Poisson models. The idea behind these models is that the excess zeros are modeled separately. With a given probability an observation is a zero. With one minus that probability it is an observation with a positive number. The probability of the zero typically follows a binary logit model using the explanatory variables present in the rest of the model. For more details we refer to Chapter 6 of Winkelmann (2008).

Model estimation

In total we estimated four models that differ on the underlying statistical distribution (Negative-Binomial versus Poisson) and whether they allow for zero-inflation or not. All of the models were estimated in STATA 11.0. Table 4 summarizes model fit for all four models.

The models are compared based on the log-likelihood, the Akaike Information Criterium (AIC) and Bayesian Information Criterium (BIC). The AIC and BIC statistics balance model fit and complexity. For these measures, lower values are more preferred. The Negative-Binomial model without zero-inflation fits the data best. It is interesting to see how accounting for zero-inflation greatly improves fit under the Poisson distribution but does not lead to any improvements under the Negative-Binomial distribution. Apparently, the large amount of zeros is sufficiently captured by the overdispersion implied by the Negative-Binomial distribution.

Estimation results

Table 5 contains the parameter estimates for the (best-fitting) Negative Binomial model. In terms of content we find support for hypotheses 1-2 that more controversial topics and topics with practical utility result in more comments. However, the self-centeredness of the post (H3) and the ambiguity of the topic (H4) did not affect the amount of comments significantly. The results reflect a lack of ambiguity aversion. This finding may be explained by the fact that the countries that are most represented in the group's membership (US, UK, The Netherlands, India) score relatively low on Hofstede's (2001) uncertainty avoidance dimension (respectively 46, 35, 53, and 40 versus a world average of 64).

With respect to post characteristics, the length of the post negatively affects the amount of comments, in support of H5. However, sentence length did not have a significant effect (H6). Readability did have a significant positive effect on the amount of discussion following a post, in support of H7. The emotionality of a post and the use of jargon have no substantial effect (H8-10). Explicitly phrasing a post as a question (H11) and encouraging members to respond (H12) both increase the number of comments, while including a hyperlink reduces the number of comments significantly (H13).

Socio-economic status (SES) is the only author characteristic that significantly increases the number of comments a post receives, in support of H15. The number of connections (H14) did not have any effect. Posting in a weekend significantly reduces the number of following comments (H16). Finally, there is no evidence for the effect of the author's gender or for a trend in the data.

Next, we compare the relative strength of each characteristic. Figure 4 displays the elasticities corresponding to variables with a significant parameter estimate. The elasticities are evaluated under the average values of the variables and presented in order of their absolute magnitude. Post readability has the largest elasticity (2.62); when readability increases by 1% the expected number of comments increases by 2.62%. Content controversy and socio-economic status of the author are responsible for the second and third highest elasticity, with respectively 1.35 and 1.29. It is interesting to note that post as well as content and author characteristics represent the top three elasticities. The two characteristics that complement the top 5 are practical utility and post length. A 1% increase in practical utility respectively 1% decrease in post length results in .98% respectively .38% increase in the expected number of comments. The remaining elasticities are rather small. None of them exceeds 0.1.

Discussion

Conclusions

With the rise of social media, online discussion forums have become important channels for firms to interact with their customers. Firms in high-knowledge industries stimulate online discussion on their forum with the goal of being perceived as thought leaders and/or build their brands. Key to the forum's success is the amount of discussion that takes place on it. Our study investigates what features and characteristics affect the number of comments that a post receives on an online discussion forum. Our empirical setting involves a global manufacturer (Philips) connecting with health care professionals through a LinkedIn discussion group. We projected that (i) content characteristics, (ii) post characteristics, (iii) author characteristics, and (iv) timing characteristics of a post jointly determine the number of comments it receives. The basis for testing our conceptual framework is formed by a collection of 316 threads; i.e. a post and following comments. Using count data models we established the effects of the different types of characteristics on the number of comments. In particular, the number of comments is higher for posts that are (i) more readable (elasticity η of 2.62), (ii) are more controversial (η = 1.35),, (iii) are written by an author with higher perceived socio-economic status (η = 1.29), (iv) have higher practical utility (η = .98), and (v) are shorter (η = .38%).

We believe our methodology is a substantial advancement over industry practice of merely studying descriptive statistics. In fact, our study was the first in-depth statistical analysis of behavior of members of the focal discussion group. It has ignited a broader research agenda by the hosting firm Philips and LinkedIn. The results of our study will be used by the involved firms in an attempt to increase the amount of discussion on the group. This will be done by communicating a top 5 recommendations on how to write a discussion generating post. LinkedIn will most likely also communicate this more broadly on their main website.

Our study addresses the implications of new media platforms for marketing communications, in particular how firms can best “seed” customer-to-customer interactions – a key research priority as identified by the Marketing Science Institute (2008). Our results enable firms hosting online discussion forums to start more promising discussions, and thus increase the appeal of the forum and consequently the sponsoring firm as thought leader in the industry.

Limitations and directions for future research

In this section, we describe some limitations and possible extensions of our study. The first limitation of our study is the sample size. The use of subjective data coded by professional judges restricted the number of threads we could use. However, we believe that this is compensated for by the depth of insights.

The generalizability of our results to other settings, such as different online discussion forums or (micro)blogs is a worthy topic of future research. We believe that our methodology is appropriate for such settings, and that the importance of readability and controversy is likely to generalize across social media types. However, several other findings may differ. Specifically, our hypotheses H10 (use of jargon) and H16 (weekend versus weekday) would reverse for social media settings that focus mostly on entertainment: the use of jargon would be detrimental, and weekend posts may generate most comments. Moreover, hyperlinks (H13) may generate more comments when media richness is high. Likewise, the relative order of importance may differ; e.g. the number of connections may be more important than socio-economic status when entertainment is the social media focus, and source characteristics may not matter at all when the level of self-disclosure is low. Thus, future studies could analyze the generalizability of our findings along the dimensions of Table 1. Our study puts the emphasis on the quantity of discussion rather than the quality of discussion. Future research could look into the challenging task of operationalizing and measuring the quality of the discussion. This would probably not only depend on the characteristics of the initial post but also on those of the following comments. Especially promising would be a joint model of quality and quantity, including their interdependency.

The ultimate goal of the company running the discussion forum in our empirical application is to be perceived as thought leader. The link between membership of the discussion group and activity on the platform on the one hand and perceptions of thought leadership on the other still needs to be formally proven. The focal firm is currently undertaking a study in joint cooperation with LinkedIn to empirically test this causal relation.

The quest to determine the ROI on online engagement continues for many in marketing. To what extent do more discussions, higher-quality discussions, active versus passive behavior of members lead to an increase in relevant metrics for the firm such as brand attitude and purchase intention? Within Philips, one of the leading metrics is the Net Promotor Score (NPS; Reichfeld 2003). Currently, Philips and LinkedIn are jointly investigating how group membership and activity within the group drive NPS scores. Initial results show that membership has the ability to increase both perceptions of thought leadership as well as NPS scores. Further research will be undertaken to validate these findings. In addition, future research will look at how successful the online group is in terms of generating insights, sales leads and partnerships for innovation.

As already highlighted by Steyer, Garcia-Bardidia and Quester (2006), online discussion groups have the potential to be great sources for data collection, as the discussions can be recorded in real time and information is available regarding the source and the sequence of the messages. We hope our study inspires research into how this potential can be unlocked.

APPENDIX

In this Appendix we show how to adjust the formula for the elasticity with respect to a metric variable (cf. Equation (3)) when estimating the model with the standardized versions of the metric variables. Suppose we include the metric variable j, x_{ij}, in its standardized form. That is, we include z_{ij} instead, which is given by:

$$z_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j}, \tag{A1}$$

where μ_j and σ_j are respectively the average and standard deviation of the j^{th} variable across posts. Suppose that $\underline{\beta}_j$ now refers to the coefficient corresponding to the standardized variable z_{ij} instead of the unstandardized variable x_{ij}. We can now rewrite the elasticity in terms of the coefficient of the standardized variable as follows:

$$\eta_{x_{ij}}^{\lambda_i} = \frac{\partial \lambda_i}{\partial x_{ij}} \cdot \frac{x_{ij}}{\lambda_i} = \left(\frac{\partial \lambda_i}{\partial z_{ij}} \cdot \frac{\partial z_{ij}}{\partial x_{ij}} \right) \cdot \left(\frac{z_{ij}}{\lambda_i} \cdot \frac{x_{ij}}{z_{ij}} \right) = \underbrace{\left(\frac{\partial \lambda_i}{\partial z_{ij}} \cdot \frac{z_{ij}}{\lambda_i} \right)}_{\substack{=\eta_{z_{ij}}^{\lambda_i} \text{ according} \\ \text{to Equation (3)}}} \cdot \left(\frac{\partial z_{ij}}{\partial x_{ij}} \cdot \frac{x_{ij}}{z_{ij}} \right) = \underline{\beta}_j z_{ij} \cdot \left(\frac{\partial z_{ij}}{\partial x_{ij}} \cdot \frac{x_{ij}}{z_{ij}} \right) =$$

$$= \underline{\beta}_j x_{ij} \cdot \frac{\partial z_{ij}}{\partial x_{ij}} = \frac{\underline{\beta}_j x_{ij}}{\sigma_j} \tag{A2}$$

References

Aggarwal, R., Gopal, R., Sankaranarayanan, R., and Singh, P. V. (2010), "Blog, blogger and the firm: Can negative employee posts lead to positive outcomes?" Forthcoming in *Information Systems Research.*

Ardichvili, A., Page, V., and Wentling, T. (2003), "Motivation and barriers to participation in virtual knowledge-sharing communities of practice," *Journal of Knowledge Management,* 7 (1)*,* 64-77.

Beenen, G., Ling, K., Wang, X., Chang, K., Frankowski, D., Resnick, P., and Kraut, R. E. (2004), "Using social psychology to motivate contributions to online communities," *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work*, 212-221, Chicago, Illinois.

Berger, J., and Milkman, K. (2012), "What makes online content viral?," Forthcoming at the *Journal of Marketing Research*.

Bickart, B., and Schindler, R. M. (2001), "Internet forums as influential sources of consumer information," *Journal of Interactive Marketing,* 15 (3)*,* 31-40.

Bishop, J. (2007), "Increasing participation in online communities: A framework for human-computer interaction," *Computers in Human Behaviors*, 23 (4)*,* 1881-1893.

Bloomberg (2011), "LinkedIn passes Myspace to become no. 2 U.S. social network," Retrieved October 3, 2011 from http://www.bloomberg.com/news/2011-07-08/linkedin-tops-myspace-to-become-second-largest-u-s-social-networking-site.html.

Boyd, D., and Ellison, N. B. (2007), "Social network sites: Definition, history, and scholarship," *Journal of Computer-Mediated Communication*, 13 (1)*,* 210-230.

Chen, Y., and Xie, J. (2008), "Online consumer review: Word-of-mouth as a new element of marketing communication mix," *Management Science,* 54 (3)*,* 477-491.

Cheung, C., Lee, M. K. O, and Rabjohn, N. (2008), "The impact of electronic word-of-mouth. *Internet Research*," 18 (3), 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 (6), 345-354.

Cialdini, R. B., Borden, R. J., Thorne, A., Walker, M. R., Freeman, S., Sloan, L. R. (1976), "Basking in reflected glory: Three (football) field studies," *Journal of Personality and Social Psychology,* 34 (3), 366-375.

Cohen, J. (1988), *Statistical power analysis for the behavioral sciences* (2nd ed.)*,* Hillsdale, NJ: Lawrence Erlbaum.

Daft, R. L., and Lengel, R. H. (1986), "Organizational information requirements, media richness and structural design," *Management Science,* 32 (5), 554-571.

Dellarocas, C. (2003), "The digitization of word-of-mouth: Promise and challenges of online reputation systems," *Management Science*, 49 (10), 1407-1424.

De Oliveira, S. M. (2003), "Breaking conversational norms on a Portuguese users network: Men as adjudicators of politeness?," *Journal of Computer-Mediated Communication,* 9 (1), Retrieved October 3, 2011 from http://jcmc.indiana.edu/vol9/issue1/oliveira.html.

Dholakia, U. M., Bagozzi, R, P., and Pearo, L. K. (2004), "A social influence model of consumer participation in network - and small-group based virtual communities," *International Journal of Research in Marketing,* 21 (3), 241-263.

Festinger, L. (1957), *A theory of cognitive dissonance,* Stanford, CA: Stanford University Press.

Fiske, S. T. (2004), *Social beings: A core motives approach to social psychology,* John Wiley and Sons, Hoboken, NJ.

Godes, D., and Mayzlin, D. (2004), "Using online conversations to study word-of-mouth communication," *Marketing Science*, 23 (4), 545-560.

---, and --- (2009), "Firm-created word-of-mouth communication: A field-based quasi-experiment," *Marketing Science*, 28 (4), 721-739.

Grice, H. P. (1975), "Logic and conversation. In P. Cole and J.J. Morgan," eds. *Syntax and Semantics, vol.3: speech acts.* New York: Academic Press.

Gunawardena, C. N., Lowe, C. A., and Anderson, T. (1997), "Analysis of a global online debate and the development of an interaction analysis model for examining social construction of knowledge in computer conferencing," *Journal of Educational Computing Research*, 17 (4), 397-431.

Hennig-Thurau, T., Gwinner, K. P., Walsh, G., and Gremler, D. D. (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.

Hoffman, D. L., and Fodor, M. (2010), "Can you measure the ROI of your social media marketing?," *MIT Sloan Management Review*, 52 (1), 41-49.

Hofstede, G. (2001), Culture's consequences, comparing values, behaviors, institutions, and organizations across nations thousands oaks, CA: Sage Publications.

Java, A., Song, X., Finn, T., and Tseng, B. (2007), "Why we Twitter: Understanding microblogging usage and communities," Proceedings of the Joint 9th WEBKDD and 1st SNA-KDD Workshop, 2007.

Johnson, E., and Payne, J. W. (1985), “Effort and accuracy in choice,” *Management Science*. 31 (4), 395-414.

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

Lee, A., Aaker, J., and Gardner, W. (2000), “The pleasures and pains of distinct self-construals: The role of interdependence in regulatory focus,” *Journal of Personality and Social Psychology*, 78 (June), 1122-1134.

Libai, B., Muller, E., and Peres, R. (2009), “The social value of word-of-mouth programs: acceleration versus acquisition,” Working paper.

LinkedIn (2010a), “Energy 360 on LinkedIn: British Gas for Business guides the energy efficiency debate at work,” Retrieved November 21, 2011 from http://marketing.linkedin.com/success-stories.

--- (2010b), “LinkedIn European business awards: New awards programme places LinkedIn and Cisco WebEx as leaders of global online collaboration,” Retrieved November 21, 2011 from http://marketing.linkedin.com/success-stories.

--- (2010c), “Business brains: Sage partners with LinkedIn on ‘Train Your Business Brain’ campaign,” Retrieved November 21, 2011 from http://marketing.linkedin.com/success-stories.

--- (2011a), “Hewlett-Packard case study: Creating 2,000+ brand advocates in two weeks with LinkedIn recommendation ads,” Retrieved November 21, 2011 from http://marketing.linkedin.com/success-stories.

--- (2011b), “Philips case study: Philips establishes industry renowned healthcare and lighting communities on LinkedIn,” Retrieved November 21, 2011 from http://marketing.linkedin.com/success-stories.

--- (2012), “LinkedIn Announces Fourth Quarter and 2011 Fiscal Year Financial Results,” Retrieved February 11, 2012 from http://press.linkedin.com/node/1104.

Marketing Science Institute (2008). Research priorities 2008-2010. Cambridge, MA.

Nowson, S., Oberlander, J., and Gill, A. (2005), “Weblogs, genres and individual differences,” *Proceedings of the 27th Annual Conference of the Cognitive Science Society*, Hillsdale, NJ, Lawrence Erlbaum Associates, 1666-1676.

Philips (2011), “*Innovations* In Health Factsheet,” October 2011.

Reichfeld, F. F. (2003), “The one number you need to grow,” *Harvard Business Review*. December. 1-11.

Schler, J., Koppel, M., Argamon, S., and Pennebaker, J. (2006), “Effects of age and gender on blogging,” *Proceedings of the AAAI Spring Symposium Computational Approaches to Analying Weblogs*, Menlo Park, CA, 191-197.

Stephen, A., and Lehmann, D. R. (2009a), “Is anyone listening? Modeling the impact of word-of-mouth at the individual level,” Working paper, Columbia University.

---, and --- (2009b), “Why do people transmit word-of-mouth? The effect of recipient and relationship characteristics on transmission behaviors,” Working paper, Columbia University.

---, and Toubia, O. (2010), “Deriving value from social commerce networks,” *Journal of Marketing Research*, 47 (2), 215-228.

Steyer, A., Garcia-Bardidia, R., and Quester, P. (2006), “Online discussion groups as social networks: An empirical investigation of word-of-mouth on the internet,” *Journal of Interactive Advertising*, 6 (2), 45-52.

The CMO Survey (2011). cmosurvey.org. August 2011, Highlights and insights, Table 5.1.

Trusov, M., Bucklin, R. E., and Pauwels, K. H. (2009), “Effects of word-of-mouth versus traditional marketing: Findings from an Internet social networking site,” *Journal of Marketing*, 73 (5), 90-102.

---, ---, and --- (2010), “Monetary value of word-of-mouth marketing in online communities,” Forthcoming in *Gfk-Marketing Intelligence Review*, Vol. 2 No. 1.

Washington, S. P. , Karlaftis, M. G., Mannering, F. L. (2003), *Statistical and econometric methods for transportation data analysis*, Boca Raton: Chapman & Hall, Boca Raton, Florida.

Wasko, M. McLure, and Faraj, S. (2005), “Why should I share? Examining social capital and knowledge contribution in electronic networks of practice,” *MIS Quarterly*, 29 (1), 35-58.

Wiertz, C., and de Ruyter, K. (2007), “Beyond the call of Duty: Why Consumers Contribute to Firm-Hosted Commercial Online Communities,” *Organization Studies*, 28 (3), 347-376.

Winkelmann, R. (2008). *Econometric Analysis of Count Data*. Springer-Verlag, Fifth Edition.

TABLE 1
OVERVIEW OF SOCIAL MEDIA TYPES AND CHARACTERISTICS

Social media type	Level of self-disclosure	Informative vs. entertaining	Personal page	Post frequency per user	Media richness
Collaboration	low	informative	no	low	low
Online gaming	low	entertaining	no	medium	high
Multi-media uploading	low	mostly entertaining	yes[#]	low	high
Weblog	high	both	yes	medium	low
Microblog	high	both	yes	high	low
Social networking site	high	mostly entertaining	yes	medium	high
Discussion group	low	mostly informative	no	medium	low

Note. The dotted lines illustrate that our empirical application, an online discussion forum for healthcare professionals on LinkedIn, can be seen as a hybrid between a discussion group and a social networking site.
[#] There are also multi-media uploading sites that do not have personal pages.

TABLE 2
VARIABLE DESCRIPTION

Variable	Definition
$CONTR_i$	the perceived controversiality of post *i*
$PRACT_i$	the perceived practical utility of post *i*
$SELF_i$	the perceived self-centeredness of post *i*
$AMBIG_i$	the perceived topic ambiguity of post *i*
$POST\_LENGTH_i$	the length of post *i*
$SENT\_LENGTH_i$	the average sentence length of post *i*
$READ_i$	the perceived readability of post *i*
POS_i	the amount of positive information contained in post *i*
NEG_i	the amount of negative information contained in post *i*
$JARGON_i$	the perceived use of jargon of post *i*
$QUESTION_i$	the perceived degree to which jargon is used in post *i*
$ENCOUR_i$	1 if the authors of post *i* encourages readers to comment, 0 otherwise
$HYPER_i$	1 if post *i* contains a hyper link, 0 otherwise
$NUM\_CONNECT_i$	the number of LinkedIn connections of the author of post *i*
SES_i	the perceived socio-economic status of the author of post *i*
$WEEKEND_i$	1 if post *i* was posted in a weekend (Saturday or Sunday), 0 otherwise
$FEMALE_i$	1 if the author of post *i* is female, 0 otherwise
$TREND_i$	monthly trend value for post *i*

TABLE 3
DESCRIPTIVE STATISTICS

Variable	Average	Std.	Min.	Max.
Number of comments	3.67	19.27	0.00	312
Content characteristics				
Controversy	1.88	1.29	1.00	7.00
Practical utility	2.31	1.25	1.00	7.00
Self-centeredness	3.41	2.04	1.00	7.00
Topic ambiguity	3.13	1.27	1.00	7.00
Post characteristics				
Post length	134.50	127.95	8.00	629.00
Sentence length	14.73	6.91	2.00	73.00
Readability	3.74	1.00	1.00	6.00
Positivity	3.23	2.81	0.00	16.67
Negativity	.72	1.67	0.00	12.50
Use of jargon	2.38	1.32	1.00	7.00
Question	.43	.50	0.00	1.00
Encouragement	.12	.33	0.00	1.00
Hyperlink	.62	.49	0.00	1.00
Author characteristics				
Number of connections	341.69	419	0.00	2,424.00
Socio-economic status	4.50	1.25	1.00	7.00
Timing characteristic				
Weekend	.16	.37	0.00	1.00
Control variable				
Author gender (female)	.29	.45	0.00	1.00

TABLE 4
OVERVIEW OF MODEL FIT

Model	Fit statistic		
	Log-likelihood	**AIC**	**BIC**
Poisson	-795	1,628	1,699
Negative-Binomial	**-424**	**888**	**963**
Zero-inflated Poisson	-722	1,488	1,571
Zero-inflated Negative-Binomial	-424	892	974

Note. In bold the best fitting model.

TABLE 5
PARAMETER ESTIMATES FOR NEGATIVE BINOMIAL MODEL

Parameter	Hypothesized sign	Coefficient	Standard error	Z-value	P-value
Intercept	N.A.	-.17	.30	-.56	.57
Content characteristics					
Controversy	+	**.93**	.09	10.04	.00
Practical utility	+	**.53**	.10	5.18	.00
Self-centeredness	–	.10	.16	.61	.54
Topic ambiguity	–	-.20	.13	-1.56	.12
Post characteristics					
Post length	–	**-.36**	.14	-2.56	.01
Sentence length	–	-.06	.12	-.46	.64
Readability	+	**.70**	.14	4.95	.00
Positivity	+	.03	.11	.26	.79
Negativity	–	-.08	.08	-1.00	.32
Use of jargon	+	.09	.09	.92	.36
Question	+	**.69**	.25	2.80	.01
Encouragement	+	**.69**	.26	2.67	.01
Hyperlink	–	**-.81**	.25	-3.23	.00
Author characteristics					
Number of connections	+	-.05	.16	-.30	.77
Socio-economic status	+	**.36**	.10	3.56	.00
Timing characteristic					
Weekend	–	**-.64**	.32	-1.98	.05
Control variables					
Author gender (female)	N.A.	.10	.24	.41	.68
Monthly trend	N.A.	-.04	.04	-1.02	.31

Note. In bold the parameters that are significant at 95%.

FIGURE 1

SCREEN SHOT OF THE INNOVATIONS IN HEALTH GROUP

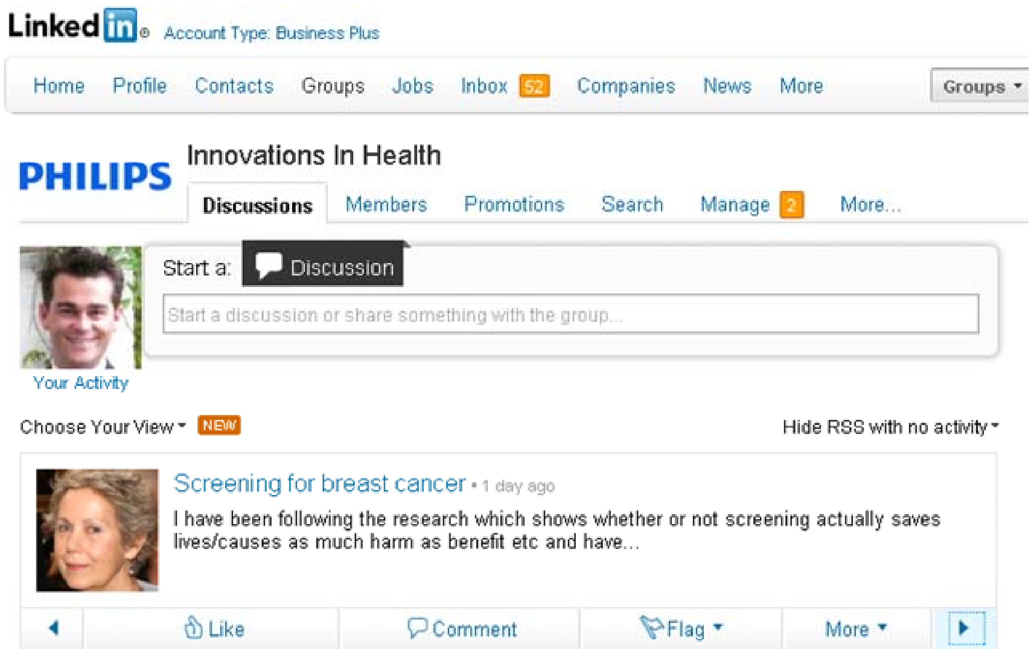


FIGURE 2
CONCEPTUAL FRAMEWORK

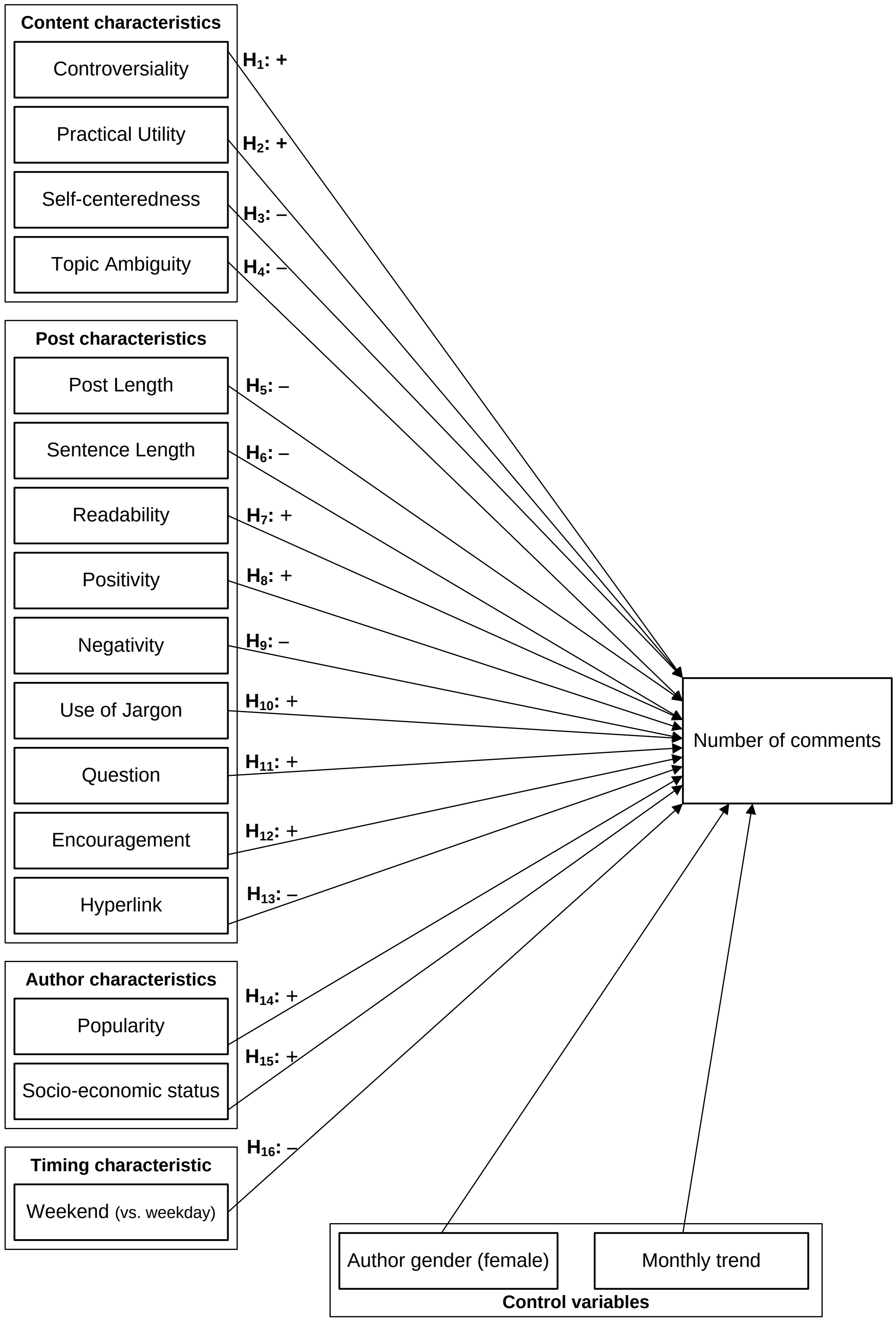


FIGURE 3
DISTRIBUTION OF NUMBER OF COMMENTS

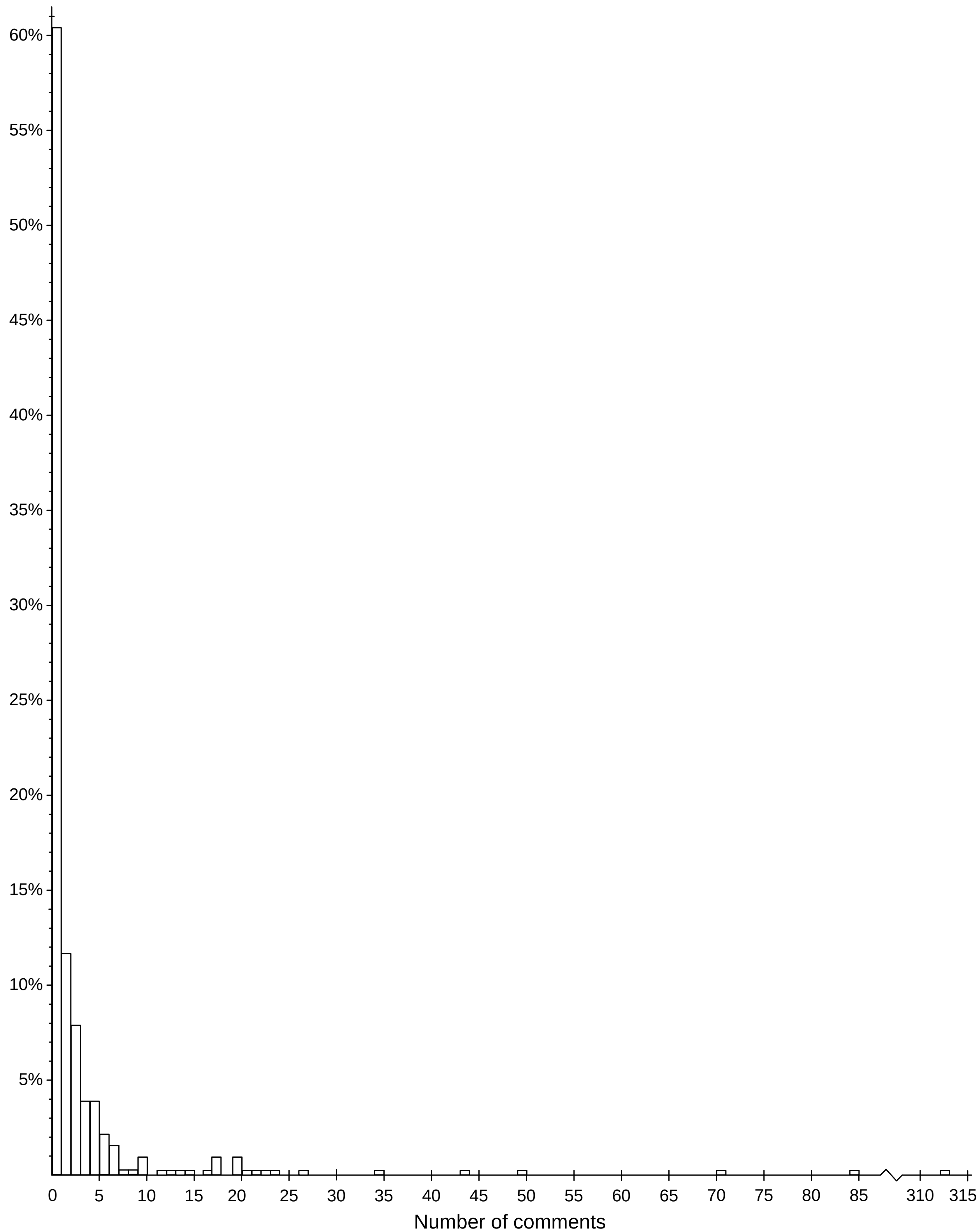
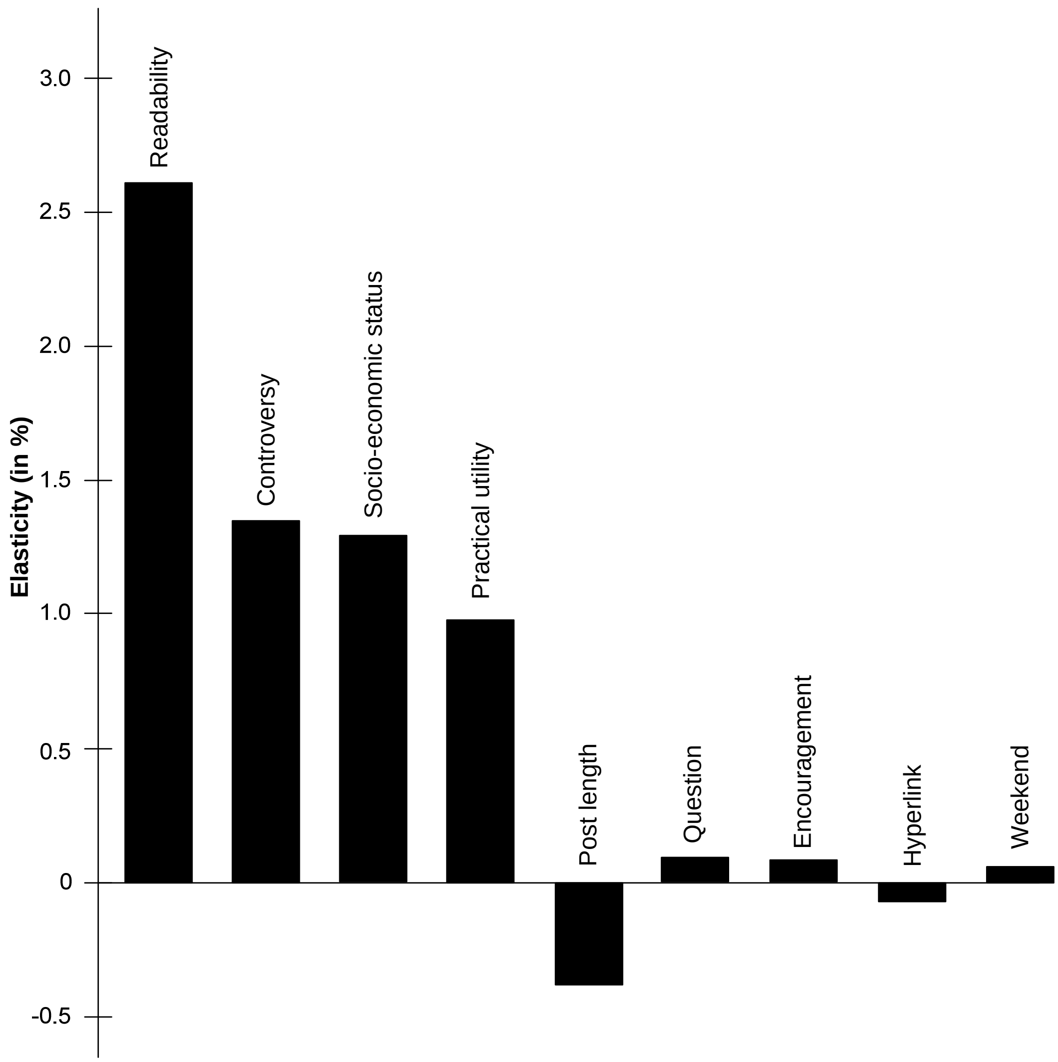


FIGURE 4
OVERVIEW OF ELASTICITIES



Note. Only elasticities pertaining to significant parameters are depicted. The elasticities are evaluated for the average value of the corresponding variable. The order of depiction is based on their absolute magnitude.