Dimensions of Leadership and Social Influence in Online Communities

David Huffaker

Media, Technology and Society, Northwestern University, Evanston, IL 60208, USA

The purpose of this article is to examine the communication behaviors of online leaders, or those who influence other members of online communities in triggering message replies, sparking conversation, and diffusing language. It relies on 632,622 messages from 33,450 participants across 16 discussion groups from Google Groups that took place over a 2-year period. It utilizes automated text analysis, social network analysis, and hierarchical linear modeling to uncover the language and social behavior of online leaders. The findings show that online leaders influence others through high communication activity, credibility, network centrality, and the use of affective, assertive, and linguistic diversity in their online messages.

doi:10.1111/j.1468-2958.2010.01390.x

Online communities, in which large groups of Internet users with common interests or activities communicate and share resources with one another (Preece & Maloney-Krichmar, 2003), constitute a considerable portion of Internet use (Horrigan, 2001). They formed early on through Internet technologies such as electronic mailing lists, chat rooms, and multiuser dungeons (Baym, 2000; Rheingold, 2000; Turkle, 1995) and more recently manifest through social networking services such as Facebook or MySpace, content-sharing Websites such as YouTube or Flickr, online discussion groups such as Google Groups, and a variety of message boards. Even blogs, inundated with comments from loyal readers or first-time viewers, represent a thriving and influential online community.

The members of these online groups create and share information at an unprecedented level, resulting in millions of messages, photos, or videos, but more importantly opinions, ideas, and a finger on the pulse of the needs and beliefs of the massive audience that makes up the Internet. Although the study of online discussion groups became an early focal point for researchers (Sudweeks, McLaughlin, & Rafaeli, 1998), we know little about characteristics of individual users who spark communication or influence its content. A deep understanding of the communication behaviors, social
networks, and language of online influentials provides a theoretical foundation for understanding social influence and information exchange on the Internet.

This article asks: What are the primary communication traits, including both linguistic characteristics and social interaction patterns, associated with those who trigger feedback and shape conversation? The purpose of this research is to make descriptive and theoretical contributions to our understanding of the ways in which Internet users are able to lead online groups by instigating and facilitating communication. Understanding the communication behaviors of influentials not only helps explain how communities thrive from user-generated content but also provides insights regarding leadership and social influence in offline settings, including businesses organizations and educational institutions. This study also aims to make a methodological contribution by demonstrating the utility of text analysis and social network analysis in understanding online behavior. By combining the study of language and social networks, scholars can produce a deeper analysis of online interaction. In effect, this article fills an important void in computer-mediated communication by providing a foundation for measuring leadership and social influence in large-scale online networks.

Background

Although leaders in the offline world have been conceptualized in many ways—as appointed or elected, as CEOs or spiritual leaders, as charismatic or transformational (Bass, 1990; Bass & Avolio, 1994)—this study conceptualizes leaders as those who stoke the fires of online communities by encouraging communication and social interaction. It draws largely on research in social influence, which suggests that leaders are best defined in terms of followers and their ability to influence attitudes and behavior (Hollander, 1961), including the ability to direct or coordinate the efforts and behaviors of an entire group (Turner, 1991).

This definition of leadership harkens to the rich history on opinion leadership in communication studies. The original definition of an opinion leader is one who could relay messages from mass media to local personal networks (Katz & Lazarsfeld, 1955), which would inevitably spread innovation throughout a community (Rogers, 1995). More recent work identifies opinion leaders as those who diffuse information or advice by discussing issues with other people in various forums with the hopes of shaping opinions (Weimann, 1994). Some scholars have described opinion leaders as gatekeepers of information, in which other members of a group actively seek out information or advice (Burt, 1999). In both cases, the relationship between leaders and their followers is clear—leaders emerge from their ability to attract followers and create a reaction.

Therefore, leaders are defined here as those who have the ability to trigger feedback, spark conversations within the community, or even shape the way that other members of a group “talk” about a topic. In other words, online leaders are those who can set agendas by causing or facilitating dialog focused on a particular topic, or frame discussion by shaping the way a particular topic is talked about. For example,
an online leader would be more likely to create a discussion about a political issue, or frame the way it is talked about in terms of the words used (i.e., republigoon or demotard). Their role is an important one; as Kerr (1986) argues, online groups are most successful when a leader is present, setting agendas, moderating interactions, and keeping the group on track with its goals.

Online leaders are defined as those who are the most capable of influencing other members of the community, and this article argues that they do so by utilizing a common set of attributes outlined by scholars in communication studies and social psychology. These include source factors (i.e., attributes routinely associated with the individual contributor such as credibility) and aspects of the actual content of the communication (O’Keefe, 2002). In computer-mediated communication, source factors include the communication behaviors of users such as how long they have been a part of the community, or how often they contribute. The message content that users contribute not only includes the subject of the individual messages but also the emotional valence associated with a message, or the way in which it is framed. This study examines each of these dimensions and their impact on the rest of the community.

**Communication activity and online leadership**

The first factor associated with leadership and social influence is simply sociability or gregariousness. In general, individuals use sociability in group settings in order to get information and build relationships (Rice, 1987). Engaging in more communication activity raises the potential to influence and extends the reach of the individual (Weimann, 1994). There are similar findings online; several studies in computer-mediated communication show that group members often perceive leaders based on the volume of their communication and follow accordingly (Misiolek & Heckman, 2005; Sudweeks & Simoff, 2005; Yoo & Alavi, 2004). Butler (2001) argues that communication activity creates online social structures that facilitate information exchange, influences social behavior, and even draws new users into the fold. Scholars also argue that leaders in successful online communities spend time motivating participation from other members in order to foster a sense of social identity within the group (Koh, Kim, Butler, & Bock, 2007). Therefore, one expects to see a relationship between online communication activity and an individual’s ability to influence the group by sparking online dialog and shaping discussions:

**H1.** The more frequently one posts messages to the group, (a) the more message replies he or she will receive from group members, (b) the longer the chain of conversation that follows his or her messages will become, and (c) the more words that he or she uses will be repeated by other group members in subsequent replies.

**H2.** The more frequently one replies to other group members, (a) the more message replies he or she will receive from group members, (b) the longer the chain of conversation that follows his or her messages will become, and (c) the more words that he or she uses will be repeated by other group members in subsequent replies.
Scholars have also noted that the credibility of the individual is very important for fostering trust and increasing influence (Pornpitakpan, 2004). Empirical work on persuasion and social influence often conceptualizes credibility in terms of expertise or trustworthiness (O’Keefe, 2002). On the Internet, reputation systems were developed early on in order to build trust between strangers and serve as a marker of credibility. For example, Ebay and Amazon have rating systems for buyers and sellers, whereas popular message boards and discussion groups often showcase an individual’s contributions to, and length of membership in, the community as a way to establish reputation and credibility. Flanagin (2007) argues that reputation systems are necessary to reduce uncertainty that occurs when individuals engage in new online relationships.

Trustworthiness can be built through the length of time spent within a group. For instance, Hollander (1961) argues that individuals can only attain leadership or influential status when they are in a group long enough for others to recognize their contribution to the group’s goals or purpose. As individuals demonstrate their dedication to an online community through participation and contribution, they accrue social trust and a positive reputation, which leads to more coordination and cooperation among members and ultimately the success of the online community (Preece, 2000). Therefore, it is expected that individuals who stimulate communication and yield influence will represent more credibility in terms of the length of time they spend in online groups:

H3. The longer the time period between one’s earliest and most recent messages, (a) the more message replies he or she will receive from group members, (b) the longer the chain of conversation that follows his or her messages will become, and (c) the more words that he or she uses will be repeated by other group members in subsequent replies.

Social networks and online leadership

Many scholars agree that there is a positive relationship between Internet participation and the creation and maintenance of social capital (DiMaggio, Hargittai, Neuman, & Robinson, 2001), which Putnam (2000) defines simply as “connections among individuals” (p. 19). For example, Ellison, Steinfield, and Lampe (2007) find a positive association between participation on Facebook and accruing social capital, primarily intended to enhance existing offline relationships. However, it is not clear how these social network behaviors impact communication activity within the group.

Studies of organizational behavior show that influential leaders tend to occupy more central positions in their surrounding networks, meaning that they tend to be highly connected with many members of the community or organization. This central role serves to increase their status, amplify their reputation, and contributes to their ability to influence others (Mehra, Dixon, Brass, & Robertson, 2006). As Weimann (1994) notes, influentials inhabit many social networks, have many friends and acquaintances, and simply “come into contact with many people” (p. 78). This increases social capital by encouraging inclusivity and reciprocity (i.e., quid pro quo) (Putnam, 2000). If central network positions offer the same rewards in online
communities, then we would expect to see individuals who drive communication to demonstrate more connections with others, including those of a reciprocal nature. While McClure Wasko, Teigland, and Faraj (2009) show that online community users do not show a “tit-for-tat” reciprocity in terms of answering specific questions, others argue that online group settings encourage reciprocal exchanges in order to navigate the sheer amount of information available (Rice, 1987).

Scholars have also noted how acting as a bridge to otherwise weakly connected subgroups can increase one’s influence. For example, Burt’s (1992) concept of “structural holes” suggests that individuals enjoy a strategic competitive advantage when they act as a broker between weakly connected and disconnected actors or subgroups. The brokering position allows individuals not only to connect the gaps between groups but also to disseminate and regulate information at their will. Having the power to sculpt and spread new information creates a powerful and influential position for brokers. Although this is especially relevant in organizational settings, Pentland, Choudhury, Eagle, and Singh (2005) state that individuals who occupy brokering positions in informal communities and social groups also demonstrate more influence.

In terms of online communities, the research findings are mixed. Ganley and Lampe (2009) find that brokering is especially popular—and beneficial—for newcomers to an online community such as Slashdot, a news aggregating site, but less so for those who achieve high “Karma” ratings (i.e., the reputation mechanism). Other studies of discussion groups find no connection between brokering and overall participation (Toral, Martinez-Torres, Barrero, & Cortes, 2009), or that online communities reflect generalized exchange (McLure Wasko et al., 2009) (i.e., interactions and reciprocal exchanges occur between members of a community as opposed to between two individual members; see Ekeh (1974) for details). By contrast, Bodendorf and Kaiser (2009) find that opinion leaders in online forums demonstrate high brokering scores. Likewise, in a study of an online learning community, Waters (2008) finds that the majority of “thought-leaders” demonstrate the highest levels of brokering (p. 343). Although the empirical evidence of brokering in online communities is mixed, there is still strong overall support for the benefits of brokering for social influence. Therefore, more brokering is expected among online users who create the most communication feedback and influence:

H4. The more often one communicates with different members, (a) the more message replies he or she will receive from group members, (b) the longer the chain of conversation that follows his or her messages will become, and (c) the more words that he or she uses will be repeated by other group members in subsequent replies.

H5. The more often one has reciprocal exchanges with group members, (a) the longer the chain of conversation that follows his or her messages will become, and (b) the more words that he or she uses will be repeated by other group members in subsequent replies.

H6. The more frequently one communicates with two members that do not communicate with each other, (a) the more message replies he or she will receive from
group members, (b) the longer the chain of conversation that follows his or her messages will become, and (c) the more words that he or she uses will be repeated by other group members in subsequent replies.

**Language use and online leadership**

Scholars have shown that particular linguistic choices in a message can increase social influence. Most research focuses on message clarity, powerful and powerless language, and language intensity (Ng & Bradac, 1993). Message clarity, which includes the complexity of the message, has been studied in terms of lexical diversity or “vocabulary richness” (Hosman, 2002, p. 374). As Bradac, Konsky, and Davies (1976) argue, “low diversity negatively affects listener’s judgments of a speaker’s linguistic, intellectual, and communicative ability, as well as judgments of his emotional state and social status” (p. 77). Likewise, “nonfluencies” in the written or spoken messages tend to score lower on perceived expertise; these include things like “vocalized pauses . . . superfluous repetition of words or sounds, corrections of slips of the tongue, [or] articulation difficulties” (O’Keefe, 2002, p. 185). In other words, poor language ability or linguistic diversity negatively impacts credibility and influence.

Again, previous work suggests that sociability remains an important aspect of social influence (Weimann, 1994), which implies that frequent or longer messages should generate more conversation. However, messages also need to be well written and lexically diverse in order to improve the credibility of the author. Therefore, more sociability and linguistic diversity can be expected in the messages that influentials contribute.

Similarly, research shows that the use of powerful and powerless language affects the perception of the source (Ng & Bradac, 1993). Powerful language is defined by its lack of powerless cues such as the use of hedges (e.g., “sort of,” “maybe,” tag questions (“isn’t it?”), hesitations (e.g., “um”), intensifiers (e.g., “really”), or fragmented sentences (Holtgraves & Lasky, 1999). Most findings suggest that powerful language is more persuasive or influential than powerless language (Burrell & Koper, 1998), so it is expected that online influentials will use more certainty, confidence, and assertiveness in their online messages.

A third area of research on the effects of message content on influence involves language intensity (Ng & Bradac, 1993). Hamilton and Hunter (1998) define language intensity as a “stylistic feature of language that is conveyed through the properties of emotionality . . . ranging from mild to intense” (p. 100). Scholars note that general emotional appeals are effective persuasive devices. As Forgas (2006) notes, “affect appears to influence what we notice, what we learn, what we remember, and ultimately the kinds of judgments and decisions we make” (p. 273). As message recipients process message content, positive and negative framing can be more influential because it grabs the attention of, and becomes more salient to, a recipient (Smith & Petty, 1996). In studies of online interactions, positive affect in messages reinforces a sense of community and encourages continued participation (Joyce & Kraut, 2006), whereas negative affect can result in feedback through hostile and insulting interactions (i.e.,
“flaming”). Therefore, it is expected that using more affect—both positive and negative—will increase message feedback and language diffusion. Berthold, Sudweeks, Newton, and Coyne (1998) find that messages with more original text, statements of facts (and not questions), relevant subject lines, and that address other users in the text all increase the chance it will receive a reply. These items highlight how language clarity, diversity, and intensity contribute to increased feedback and diffusion:

H7. The more often one contributes longer messages to the group, (a) the more message replies he or she will receive from group members, (b) the longer the chain of conversation that follows his or her messages will become, and (c) the more words that he or she uses will be repeated by other group members in subsequent replies.

H8. The more often one uses unique words in their online messages, (a) the more message replies he or she will receive from group members, (b) the longer the chain of conversation that follows his or her messages will become, and (c) the more words that he or she uses will be repeated by other group members in subsequent replies.

H9. The more often one uses words that express certainty or assertiveness, (a) the more message replies he or she will receive from group members, (b) the longer the chain of conversation that follows his or her messages will become, and (c) the more words that he or she uses will be repeated by other group members in subsequent replies.

H10. The more often one uses words that express affect, (a) the more message replies he or she will receive from group members, (b) the longer the chain of conversation that follows his or her messages will become, and (c) the more words that he or she uses will be repeated by other group members in subsequent replies.

Methodology

Sample
The sample consists of 33,540 users who contributed 632,622 messages to 16 different discussion groups found on Google Groups between June 21, 2003, and January 31, 2005. Google Groups allows users to post messages in a public forum and to reply to the messages of other users, which often form chronological conversational threads. These messages are typical of the type seen in message boards or discussion lists, and include question-and-answer and supportive exchanges, and discussion of news and topical issues, or expression of opinions.

Google Groups is the current manifestation of Usenet, one of the first online discussion forums. Microsoft’s Netscan project (Smith, 2007) captured all social interactions that took place on Usenet over several years, including the complete message threads, author IDs, and time/date stamps, resulting in data for thousands of groups. However, Netscan did not include the actual message content, and researchers at Carnegie Mellon University (CMU) later captured the original archived message content from Google Groups for a random subset of 99 groups from the total Netscan
data, ranging from June 2003 to January 2005 (Joyce & Kraut, 2006). The sample used in this study combines both the social interactions from the Netscan data with the message content from the CMU project. Because of the cross-sectional method of data extraction, this study does not follow these groups from their origin (some groups might be several years old).

Online discussion groups focus on a variety of topics. This sample consists of a random strata of four topic areas: (a) politics, (b) health and support, (c) recreation and hobby, and (d) science and technology, which researchers have identified as the most popular genres of online community topics (Horrigan, 2001). Therefore, these discussion groups constitute a representative sample of online communities in general. Sixteen groups were randomly sampled from these four genres using a random integer generator. They included groups focused on economics and political discussion (including radical-left and gun rights); breast cancer, hepatitis-c, and depression support groups; hobbies such as manga art, quilting, blues music, and vegetarian cooking; and groups focused on operations research, chemistry, linguistics, and computer security.

Procedure
All communication activities, including how often authors posted and replied to other messages, and the dates associated with their first and last messages during that 20-month time frame, were captured as user log files on a mySQL database. Using this database, a social graph was created for each group based on who replied to whom for all authors during the time period. In other words, if Author A (source) replied to Author B (target), this was considered a network link. These links were analyzed using UCINET, a popular social network analysis software package (Borgatti, Everett, & Freeman, 1999), to capture the network attributes presented below.

Second, the actual messages exchanged during the 20-month period were converted to text files and preprocessed to remove all headers, subject lines, signatures, and quotationed text. Each text file was then processed using Linguistic Inquiry and Word Count (LIWC) (Pennebaker, Booth, & Francis, 2006). LIWC relies on a series of built-in dictionaries that allow the classification of word frequencies into a series of rudimentary linguistic dimensions such as pronoun or verb use, as well as social and psychological processes such as positive or negative emotions. The details of LIWC are described in the measures section.

Independent measures
Measurement of communication activity
Based on the user logs, the basic communication behavior was extracted, including message posts, message replies, and the first and last time an author contributed to the community. These variables were log transformed to reduce the positive skew in the data distribution.

Number of posts. This is calculated as the total number of unique posts that an author contributes to a topic group. It does not include replies to other messages.
Number of replies. This is calculated as the total number of replies that an author contributes to messages posted by other members of a topic group.

Tenure in community. This is calculated as the total number of days during which an author actively posts a message or reply to a topic group within the sampling frame. It is calculated by subtracting the last message date from the first message date.

Measurement of social networks
In order to examine expansiveness of each participant as well as his value as an intermediary in the network, three measures of centrality outlined by Freeman (1979) and Wasserman and Faust (1994) were utilized. All three measures were log transformed.

Expansiveness. Expansiveness is measured in terms of outdegree centrality, which represents the number of outgoing links for each author. In other words, it is measured by the number of times that an author replies to different members in the community (this is different from number of replies, in which an author could reply repeatedly to only one other member).

Reciprocity. Reciprocity is measured as the frequency of an individual’s participation in a mutual dyad, in which both actors reply to one another, regardless of the order of the replies (Wasserman & Faust, 1994). In order to calculate this, the networks were converted into symmetric relations, and the sum of reciprocal links was calculated for each author. The MINIMUM function in UCINET was used to symmetrize the network, thus ensuring that a value in the matrix will only be calculated when both actors connect to one another (in other words, both authors must have responded to the other’s message in order to be included in the reciprocal analysis).

Brokering. Brokering is measured in terms of betweenness centrality, which identifies which authors link the shortest paths, or geodesics, in the network. For example, if Author B is the only node between Author A and Author C, then Author B is considered an intermediary or broker and possibly able to control the communication between them (Wasserman & Faust, 1994). The value of betweenness centrality increases as a participant resides on more and more geodesics. It is calculated by examining the sum of the proportion of geodesics linking any two actors with the same actors containing a third. Although betweenness centrality is a more general measure of brokering in terms of structural equivalence noted by Burt (1992), it is still useful for identifying intermediaries in networks (Hanneman & Riddle, 2005).

Measurement of language use
The content of the messages is also analyzed for several aspects of language using word frequency analysis. All linguistic dimensions are evaluated using predefined dictionaries outlined by Pennebaker et al. (2006). LIWC reads text files and identifies the number of times that a word matches a particular dictionary and then calculates the frequency of occurrence by the total number of words, thereby creating a ratio of
language types. For example, if a user wrote words like “love,” “good,” and “happy” (words associated with the Affect dictionary) in a message using 10 words, they get a value of 30% for that message.

Because a single author might post multiple messages, the LIWC scores represent an aggregated score for each author. In other words, after calculating the LIWC scores for each message, the data set was aggregated by individual author, using the mean score for each LIWC value. In addition, each linguistic dimension was also log transformed.

**Talkativeness.** This is measured as the average length of messages contributed by each participant, or the sum of the total words found in each message divided by the total number messages.

**Linguistic diversity.** This is measured as the number of unique words found in a message, calculated by dividing the number of different words by the total number of words. This is often referred to as a type/token ratio.

**Assertiveness.** This is measured as the frequency of certainty words such as “always” or “never” used by each participant, based on a dictionary of 83 words (http://www.liwc.net/ for dictionary details) divided by the total number of words used in each message.

**Affect.** This is measured as the frequency of words such as “happy,” “cried,” “sweet,” “nice,” or “ugly,” which represent affective or emotional language used by each participant based on a dictionary of 915 words, divided by the total number of words used in each message.

**Dependent measures**

In this study, online leadership is defined in terms of the ability to trigger communication feedback in the group, and to influence the language used in online discussion. This is operationalized in several ways: (a) as a “reply trigger,” in which users inspire others to respond to posts; (b) through “conversation creation,” or the ability to spark lengthy conversations; and (c) through “language diffusion,” or the ability to spread actual word choices. The purpose of using all three is to develop a more robust understanding of online feedback mechanisms. It should be noted that online leadership is measured on a continuous scale, rather than as discrete groups, so each hypothesis tests an association between communication activity, social network behaviors, and language use, with the ability to create more replies, conversation, or language diffusion. The purpose is to understand if using a higher frequency of certain types of language or social interaction will result in more audience feedback and language diffusion, and ultimately social influence.

**Reply trigger**

The ability to inspire users to reply to a message is measured in terms of indegree centrality, which represents the total number of incoming links (i.e., the number of responses) that a user receives (Wasserman & Faust, 1994).
Conversation creation
A second way to conceptualize online social influence occurs when users post a message or reply that sparks a long dialog between other users. For example, Author A may start a thread about a new health product and receive three direct replies. Reply 1 and 2 spur additional replies, and Reply 2 actually sparks a third-tier reply. Because Author A initiated the discussion or question, he directly or indirectly influences each of these tiers. This variable is calculated as the total number of first-tier, second-tier, third-tier, and so on replies for each author’s post. This includes the number of times that a second-tier or third-tier author also receives a reply.

Language diffusion
Social influence follows a similar pattern when language is taken into account by examining the number of shared words between two authors who interact. For example, if Author A included the term “Nike” in a post, and Author B also said “Nike” in his reply, it is considered influential. If Author C, in replying to Author B’s message, also includes “Nike,” then Author B can be considered influential along with Author A. The number of words that are repeated in subsequent replies were calculated for each individual posted message. Afterward, the number of repeated words for each individual post was summed for each individual author.

The number of repeated words is captured using Text::Similarity (Pederson, Patwardhan, Banerjee, & Michelizzi, 2008), an open-source Perl Module that captures the similarity between two files by counting the frequency of overlapping words or phrases. The frequency of shared words or phrases is then normalized by the length of each file, ensuring that longer messages do not have a greater chance of diffusing words.

In order to make sure that the words that are measured are meaningful, the messages were processed a second time to remove any stop-words (“a,” “an,” “the,” etc.). This way common words such as “the” are not considered influential and do not muddle the measure. Also, any instance where an author responds to a message in a chain that he or she started was removed in order to remove redundant scores and control for self-replicating language.

Results
Descriptive statistics
The mean, standard deviation, and correlation matrix for each variable of interest are listed in Table 1. All three dependent measures are highly correlated. Reply trigger and conversation creation are significantly correlated to all variables except talkativeness ($p = .21$ and $p = .18$, respectively) and assertiveness ($p = .65$ and $p = .56$, respectively). Language diffusion is significantly correlated to all variables except talkativeness ($p = .26$), affect ($p = .47$), assertiveness ($p = .99$), and linguistic diversity ($p = .09$).
Table 1 Correlation Matrix for All Variables of Interest

<table>
<thead>
<tr>
<th>Variable</th>
<th>M (SD)</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Reply trigger</td>
<td>14.1 (93.24)</td>
<td>.97**</td>
<td>.84**</td>
<td>.54**</td>
<td>.81**</td>
<td>.75**</td>
<td>.86**</td>
<td>.39**</td>
<td>.89**</td>
<td>-.01</td>
<td>.02**</td>
<td>.00</td>
<td>.02**</td>
</tr>
<tr>
<td>2. Conversation creation</td>
<td>12.1 (86.04)</td>
<td>—</td>
<td>.86**</td>
<td>.49**</td>
<td>.80**</td>
<td>.71**</td>
<td>.86**</td>
<td>.39**</td>
<td>.89**</td>
<td>-.01</td>
<td>.012*</td>
<td>.00</td>
<td>.02**</td>
</tr>
<tr>
<td>3. Language diffusion</td>
<td>1.1 (10.14)</td>
<td>—</td>
<td>.30**</td>
<td>.74**</td>
<td>.70**</td>
<td>.81**</td>
<td>.32**</td>
<td>.84**</td>
<td>-.01</td>
<td>.007</td>
<td>.00</td>
<td>.01</td>
<td></td>
</tr>
<tr>
<td>4. Number of posts</td>
<td>21.0 (147.00)</td>
<td>—</td>
<td>.11**</td>
<td>.23**</td>
<td>.10**</td>
<td>.27**</td>
<td>.11**</td>
<td>.01</td>
<td>.001</td>
<td>-.00</td>
<td>-.01**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Number of replies</td>
<td>18.4 (126.95)</td>
<td>—</td>
<td>.81**</td>
<td>.94**</td>
<td>.34**</td>
<td>.91**</td>
<td>-.01</td>
<td>.021</td>
<td>-.00</td>
<td>.03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Tenure</td>
<td>7.3 (26.88)</td>
<td>—</td>
<td>.78**</td>
<td>.41**</td>
<td>.73**</td>
<td>-.01</td>
<td>.018</td>
<td>-.00</td>
<td>.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Expansiveness</td>
<td>14.1 (96.24)</td>
<td>—</td>
<td>.38**</td>
<td>.97**</td>
<td>-.01*</td>
<td>.020</td>
<td>.00</td>
<td>.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Brokering</td>
<td>9,316.0 (94,329.70)</td>
<td>—</td>
<td>.30**</td>
<td>-.00</td>
<td>.003</td>
<td>.00</td>
<td>-.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Reciprocity</td>
<td>8.0 (65.46)</td>
<td>—</td>
<td>-.01</td>
<td>.018**</td>
<td>.00</td>
<td>.03**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Talkativeness</td>
<td>181.7 (752.18)</td>
<td>—</td>
<td>.003</td>
<td>.01</td>
<td>-.34**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Affect</td>
<td>5.3 (4.86)</td>
<td>—</td>
<td>.07**</td>
<td>.07**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Assertiveness</td>
<td>1.4 (2.19)</td>
<td>—</td>
<td>-.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Linguistic diversity</td>
<td>79.9 (15.09)</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Bivariate correlations, two-tailed tests. N = 33,540.
*p < .05. **p < .01.
There are small, positive correlations between the number of posts and the number of replies, and the posts and tenure in the community. There is a small correlation between expansiveness (i.e., outdegree) and brokering (i.e., betweenness) and a strong correlation between expansiveness and reciprocity. Collinearity tests suggest that both expansiveness (tolerance = .05, variance inflation factor [VIF] = 21.71) and reciprocity (tolerance = .04, VIF = 27.12) are possibly redundant variables (Hair, Anderson, Tatham, & Black, 1998). Talkativeness and linguistic diversity are negatively correlated, which resonates with previous arguments that the type-token ratio decreases as the number of words increases (Malvern, Richards, Chipere, & Durán, 2004; McKee, Malvern, & Richards, 2000). However, a collinearity test does not reveal a problem with these variables because tolerance is much greater than 0 and VIF is less than 20 (talkativeness tolerance = .83, VIF = 1.2; linguistic diversity tolerance = .83, VIF = 1.21) (see Menard, 1995 for more information on tolerance and VIF tests). There are small, but significant correlations between affect, assertiveness, and linguistic diversity. Overall, the correlation analysis suggests that aside from reciprocity and expansiveness, there are no multicollinearity issues.

Hierarchical linear modeling analysis
In order to examine the relationship between communication activity, social networks and language use, and the ability to influence the communication behaviors of others, this study relies on hierarchical linear modeling (HLM). HLM (also referred to as multilevel modeling) is often used for nested data structures and accounts for the within-group and between-group variance (Gelman & Hill, 2007). It is especially useful in this context because the authors occupy various discussion groups with their own unique characteristics (i.e., 16 different groups ranging in topic, size, and interaction level). HLM not only allows for the analysis of nested data such as these but is also considered more efficient in inferring predictor variables and more accurate in measuring the standard errors (Gelman & Hill, 2007).

Because the dependent variables represent event count data—that is, they are nonnegative and integer based—the Poisson regression model is recommended for this particular data distribution (Cameron & Trivedi, 1998). Poisson regression relies on a log transformation of the dependent variables and requires an antilog transformation of the coefficients of each predictor in the regression model to interpret the odds ratio, which is used to assess the effect size (Gardner, Mulvey, & Shaw, 1995; Gelman & Hill, 2007). Because the values of the mean and variance were similar, the analysis is adjusted for overdispersion, another recommended analytic technique (Gelman & Hill, 2007). Individual regressions were run for each dependent variable (i.e., reply trigger, conversation creation, and language diffusion) and for all independent variables (i.e., communication activity, social networks, and language use), producing three separate models.

Users who generate the most message replies, comments, or conversation, or spread the most word choices were expected to exhibit more communication activity...
Online leadership as reply trigger

Model I: Communication activity

The first hypotheses predicted that leaders are more likely to post messages (H1a) and reply to other messages (H2a) than other members of the group and that leaders are more likely to have a longer tenure (H3a) in the group. The model is a good fit, $\chi^2$ (15) = 1318.54, $p < .001$, and all three hypotheses are supported. As shown in Table 2, the number of posts, number of replies, and number of active days in the group are positively associated with the number of replies triggered by posting an online message. Note that number of posts (Exp($B$) = 3.45) demonstrates a higher odds ratio than the number of replies (Exp($B$) = 1.29). Again, odds ratios have been noted as the ideal way to measure effect size across variables (Hosmer & Lemeshow, 2004). In this case, the odds ratios (i.e., Exp($B$)) can be interpreted to mean that a one-unit change in the number of posts will produce more than three times as many replies from other group members.
Model II: Social networks
The second set of hypotheses predicts that leaders are more likely to be expansive (H4a) and more likely to serve as brokers between otherwise disconnected participants (H6a). The model is a good fit, $\chi^2 (15) = 1884.34, p < .001$, and provides partial support for the hypotheses. Although expansiveness (i.e., outdegree centrality) is positively associated with reply triggers, brokering (i.e., betweenness centrality) is not related to being a reply trigger ($p = .64$). Note that reciprocity ($\text{Exp}(B) = 6.33$) demonstrates a higher odds ratio than expansiveness ($\text{Exp}(B) = 3.43$). Because expansiveness and betweenness were correlated, additional models were created, in which each variable was dropped. The results of these models show the same magnitude, direction, and significance for each variable across all three dependent variables.

Model III: Language use
Finally, the last set of hypotheses predicted that leaders would be more likely to demonstrate higher frequencies of talkativeness (H7a), linguistic diversity (H8a), assertiveness (H9a), and affect (H10a). The model is a good fit, $\chi^2 (15) = 727.52, p < .001$, and all four hypotheses are supported. Talkativeness (measured in terms of the average message length), linguistic diversity (measured in terms of the type/token ratio), assertiveness (measured in terms of the relative frequency of certainty words in a message), and affect (measured in terms of the relative frequency of emotional words in a message) are all positively related to triggering a message reply. Note that talkativeness ($\text{Exp}(B) = 11.19$) demonstrates a higher odds ratio than affect ($\text{Exp}(B) = 1.99$) or assertiveness ($\text{Exp}(B) = 2.38$), whereas linguistic diversity is considerably higher than all three ($\text{Exp}(B) = 1154.07$).

Online leadership as conversation creation
Model I: Communication activity
The second model is also a good fit, $\chi^2 (15) = 598.45, p < .001$. As shown in Table 3, the number of posts (H1b), replies (H2b), and tenure in the community (H3b) are all positively associated with conversation creation. Contrary to the reply trigger model, the odds ratio for the number of replies ($\text{Exp}(B) = 4.63$) is higher than the number of posts ($\text{Exp}(B) = 1.17$).

Model II: Social networks
In addition to expansiveness and brokering, it was hypothesized that leaders are more likely to be reciprocal (H5a). The model is a good fit, $\chi^2 (15) = 980.66, p < .001$, and again provides partial support for the hypotheses. Expansiveness (H4b) and reciprocity (H5a) are positively associated with conversation creation, but brokering (H6b) is not related ($p = .08$). Also contrary to the reply trigger model, expansiveness ($\text{Exp}(B) = 8.94$) demonstrates a higher odds ratio than reciprocity ($\text{Exp}(B) = 3.81$).

Model III: Language use.
The model is a good fit, $\chi^2 (15) = 448.31, p < .001$, and provides support for all four hypotheses. Talkativeness (H7b), linguistic diversity (H8b), assertiveness (H9b),
Table 3  Summary of Hierarchical Linear Model for Communication Activity, Network, and Language Variables Predicting Conversation Creation in Online Communities (N = 33,540)

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>B SE</th>
<th>Exp(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>I. Communication Activity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of posts</td>
<td>0.16</td>
<td>.02</td>
<td>1.17***</td>
</tr>
<tr>
<td>Number of replies</td>
<td>1.53</td>
<td>.20</td>
<td>4.63***</td>
</tr>
<tr>
<td>Tenure in community</td>
<td>0.64</td>
<td>.21</td>
<td>1.89**</td>
</tr>
<tr>
<td><strong>II. Social networks</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expansiveness</td>
<td>2.19</td>
<td>.47</td>
<td>8.94***</td>
</tr>
<tr>
<td>Reciprocity</td>
<td>1.34</td>
<td>.16</td>
<td>3.81***</td>
</tr>
<tr>
<td>Brokering</td>
<td>-0.67</td>
<td>.36</td>
<td>0.51</td>
</tr>
<tr>
<td><strong>III. Language use</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Talkativeness</td>
<td>2.44</td>
<td>.12</td>
<td>11.51***</td>
</tr>
<tr>
<td>Linguistic diversity</td>
<td>7.63</td>
<td>.99</td>
<td>2054.02***</td>
</tr>
<tr>
<td>Assertiveness</td>
<td>0.93</td>
<td>.06</td>
<td>2.54***</td>
</tr>
<tr>
<td>Affect</td>
<td>0.68</td>
<td>.11</td>
<td>1.98***</td>
</tr>
</tbody>
</table>

*Note: The restricted maximum likelihood method is used for estimation. Predictor variables are estimates of the fixed effects, \( \gamma \)s, with robust standard errors, and adjusted for overdispersion. \( \chi^2 \) (df) for (I) = 598.46 (15);*** (II) = 980.66(15);*** and (III) = 448.31(15).***

The restricted maximum likelihood method is used for estimation. Predictor variables are estimates of the fixed effects, \( \gamma \)s, with robust standard errors, and adjusted for overdispersion. \( \chi^2 \) (df) for (I) = 598.46 (15);*** (II) = 980.66(15);*** and (III) = 448.31(15).***

and affect (H10b) are all positively associated with conversation creation. Again, talkativeness (Exp(B) = 11.51) demonstrates a higher odds ratio, whereas linguistic diversity shows a very high odds ratio (Exp(B) = 2054.02).

Online leadership as language diffusion

Model I: Communication activity

The model is a good fit, \( \chi^2(15) = 623.79, p < .001 \), and supports all three hypotheses. As shown in Table 4, the posts (H1c), replies (H2c), and tenure in the community (H3c) are positively associated with language diffusion. Again, the replies (Exp(B) = 4.12) show a higher odds ratio than posts (Exp(B) = 1.16).

Model II: Social networks

The model is a good fit, \( \chi^2(15) = 629.50, p < .001 \), and provides partial support for the hypotheses. Expansiveness (H4c) and reciprocity (H5b) are positively associated with language diffusion, whereas brokering (H6c) is not (\( p = .09 \)). In this model, reciprocity (Exp(B) = 7.46) shows a higher odds ratio than expansiveness (Exp(B) = 4.92).

Model III: Language use

The model is a good fit, \( \chi^2(15) = 542.91, p < .001 \), and supports all four hypotheses. Talkativeness (H7c), linguistic diversity (H8c), assertiveness (H9c), and affect (H10c) are all positively related to language diffusion. Again, talkativeness (Exp(B) = 14.08)
Table 4  Summary of Hierarchical Linear Model for Communication Activity, Network, and Language Variables Predicting Language Diffusion in Online Communities (N = 33,540)

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>B SE</th>
<th>Exp(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Communication Activity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of posts</td>
<td>0.15</td>
<td>.03</td>
<td>1.16***</td>
</tr>
<tr>
<td>Number of replies</td>
<td>1.41</td>
<td>.21</td>
<td>4.12***</td>
</tr>
<tr>
<td>Tenure in community</td>
<td>0.84</td>
<td>.23</td>
<td>2.31**</td>
</tr>
<tr>
<td>II. Social networks</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expansiveness</td>
<td>1.59</td>
<td>.45</td>
<td>4.92**</td>
</tr>
<tr>
<td>Reciprocity</td>
<td>2.01</td>
<td>.31</td>
<td>7.46***</td>
</tr>
<tr>
<td>Brokering</td>
<td>−0.65</td>
<td>.36</td>
<td>0.52</td>
</tr>
<tr>
<td>III. Language use</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Talkativeness</td>
<td>2.64</td>
<td>.27</td>
<td>14.08***</td>
</tr>
<tr>
<td>Linguistic diversity</td>
<td>7.57</td>
<td>1.37</td>
<td>1939.20***</td>
</tr>
<tr>
<td>Assertiveness</td>
<td>0.94</td>
<td>.05</td>
<td>2.58***</td>
</tr>
<tr>
<td>Affect</td>
<td>0.76</td>
<td>.15</td>
<td>2.14***</td>
</tr>
</tbody>
</table>

Note: The restricted maximum likelihood method is used for estimation. Predictor variables are estimates of the fixed effects, $\gamma$s, with robust standard errors, and adjusted for overdispersion. $\chi^2$ (df) for (I) = 623.79 (15);*** (II) = 629.50(15);*** and (III) = 542.91(15).*** $**p < .01. ***p < .001.$

shows a higher odds ratio than affect and assertiveness, whereas linguistic diversity is very high ($\text{Exp}(B) = 1939.20$).

**Discussion**

The purpose of this study is to examine the communication activity, social networks, and language of users who trigger replies, spark long conversations, or influence the language and word choices within Google Groups, a discussion forum that allows users to post messages and reply to others and form groups focused on a variety of topics. Those who are able to generate the most communication feedback and influence are considered online leaders because they are able to spark reaction, create conversation, and influence what is talked about in online communities by spreading word choices along a dialog.

The results show that communication activity (i.e., the number of posts and replies) and tenure within a group are indeed related to the ability to influence others. Likewise, network centrality measures such as expansiveness (i.e., the number of links to others) and reciprocity are also positively related. Interestingly, brokering—which measures the extent to which a user connects to otherwise disconnected users—was not related to the ability to encourage or influence communication. In terms of language use, online leaders show higher frequencies of talkativeness, linguistic diversity, assertiveness, and affect.
These results were consistent across all three conceptualizations of online leadership. However, the effect sizes for some predictors varied across the models. Posting new messages seems more effective for triggering a response, whereas replying to others seems more effective for creating conversations and language diffusion. Similarly, connecting to others is more effective for creating conversations, whereas reciprocity is more effective for triggering replies or language diffusion.

The finding that sheer communication activity is central to being influential is interesting because the bulk of information in most online communities is produced by a subset of users. For example, Smith (1999) finds that 18,000 people contribute 67,000 messages to Usenet daily. Wikipedia relies on a small group of editors to do the majority of work (Kittur & Kraut, 2008). There is also a substantial lurker population in most online communities, which enjoy perusing content without adding to it (Nonnecke, Andrews, & Preece, 2006). The results on this study not only confirm this rule of participation (e.g., power laws, long tail, or 80/20 rule) but also show that this subset helps drive participation in online communities. Many scholars agree that “knowledge collecting precedes knowledge contributing” in online systems; however, collecting can also lead to future contributions (Heinz & Rice, 2009, p. 154).

The importance of credibility, as outlined by social influence theorists, also plays an interesting role online. Although leaders are more likely to have longer tenure in the community, that length of time may not always be explicit to other users (unless they have also spent a lot of time in the group) watching the conversation flow or searching through post archives. However, their tenure and credibility might manifest when they demonstrate knowledge of the group’s history (e.g., “we talked about this issue six months ago”) or in their familiarity with other members (e.g., “you should talk to Steve since he knows about that problem”).

What is really interesting here is that leaders both give and receive. That is, they do not simply engage in posting behavior, but spend time replying to others. In fact, these findings show that posting behavior is outweighed by reply behavior, as least for sparking dialogs and spreading word choice. Leaders also engage in reciprocal links. That is, they tend to reply to those who reply to them. This suggests that some level of bonding or relationship development is at play in online communities, and it is a trait that leaders use to increase influence. More specifically, the results show that information is disseminated by engaging in reciprocity rather than simple broadcasting. This suggests that relationships are developed and nurtured, even in a voluntary discussion group where attrition is high. This finding reflects previous work by McClure Wasko, Teigland, and Faraj (2009), which suggests that online communities show a pattern of generalized exchange, even when there is high member turnover.

Wellman and Gulia (1999) argue that online community interactions go beyond simple information exchange and encourage supportive, loyal relationships. The findings of this study support this notion and show that leaders are often “reaching out” to the community, engaging in relationship development and maintenance that contribute to the success of the group as a whole. The result of this is more
diffusion: When people are close to others—even in terms of sharing communication messages—they are more likely to adopt an idea (Granovetter, 1978). This also resonates with Rice’s (1982) finding that reciprocity in online groups is important and that individuals must utilize both transmitting and receiving in order to foster real information exchange, or risk becoming isolated.

Contrary to expectation, brokering was not a significant predictor of online social influence. One explanation for this is that it may not be as prevalent or important in online discussion groups because all information is transparent and accessible. According to these findings, everyone in the community can serve as a broker, and leaders are not exceptional in this way. Another explanation is that brokering might have a more important role when it serves to link across separate groups, rather than connect members within them.

The results also show that talkativeness, affect, assertiveness, and linguistic diversity are linguistic traits positively associated with online leaders. The finding that online leaders tend to produce lengthier messages is interesting because previous research has suggested that computer-mediated communication lends itself to simpler messages, both in terms of vocabulary richness and utterance length (Herring, 2002). In fact, Crystal (2001) argues that this is especially true in discussion groups because the culture of Usenet imposes a “pragmatic pressure on individuals to keep their contributions relatively short” (p. 134). Therefore, despite these sociocultural conventions, leaders still contribute lengthier posts, and with a positive effect on communication.

Based on the magnitude of the coefficients, linguistic diversity has a strong relationship with leadership. Again, linguistic diversity is a measure of lexical complexity or vocabulary richness, and it is often used to measure cognitive ability (Malvern et al., 2004). In fact, listeners or readers have been shown to relate lexical diversity to the intellectual competence and communication ability of a speaker or writer (Ng & Bradac, 1993), which links directly to the importance of source credibility (O’Keefe, 2002). So it is not surprising that leaders are influential online because they are perceived as competent and credible. One might also expect that lengthy messages have a curvilinear effect on message feedback, in which very short messages do not provide enough information or effort to be worthy of reaction, whereas very long messages become too complex for readers and might be ignored because they are too difficult to understand. The linguistic diversity result implies that although leaders write longer messages, they exhibit a command of the language that keeps these messages understandable and compelling to readers.

A second way to interpret the findings is that leaders use richer, more colorful language, which draws readers in. As Thayer (1988) suggests, leadership is about telling compelling stories that enchant listeners. If the majority of messages in a topic group seem incoherent or even disorganized, it would seem likely that a reader might focus on messages that exhibit thoughtfulness and clarity. In effect, packaging information in salient, compelling, or passionate terms seems a surefire way to spark a conversation or spread an idea.
The assertiveness finding suggests that online leaders exude more confidence or certainty in their online communication and that powerful language seems a successful persuasive device that leaders utilize when trying to get a point across or debate an issue. Scholars have argued that powerful language is something prevalent in discussion groups. For example, Herring (2003) shows how aggressive behaviors prevail online, although not necessary to the benefit of all participants.

The finding that affect was related to online leadership is also important because it shows how emotional valence instigates conversation and idea flow. This confirms theories that language intensity in messages can increase persuasion (Ng & Bradac, 1993), but suggests that the role intensity plays in discussion groups might have something do with passion surrounding an issue or opinion. In fact, some have regarded discussion groups as a place where people come simply to argue and have heated debates (Kelly, Fisher, & Smith, 2005). Therefore, leaders may rely on positive and negative emotional language during deliberations to make a point and spark reaction.

There are several limitations to this research. First, trolls—who watch discussion groups and post messages to be purposely provocative, often relying on emotionally laden messages (Donath, 1998)—could be affecting the data. Provoking users can certainly result in a conversational thread (i.e., “hey get out of here troll!”); however, studies show that users often regulate troll behavior by ignoring or blocking their messages (Herring, Job-Sluder, Scheckler, & Barab, 2002). Therefore, it is unlikely that a troll would be able to survive long enough to reach high levels of feedback or diffusion. Still, future work should attempt to find patterns of communication behavior that might be used to detect and remove trolls.

A second limitation to this work is that some of the measures need refining. For example, the variable representing tenure in the community suffers from the sampling frame: Users who have inhabited a community for a long duration but stopped posting early in the time frame used here are misrepresented. However, the results still show that those with the highest total amount of time being active are the ones with the most feedback and diffusion. Still, future work should examine these models from the origin of a group. Likewise, when users all converge on a particular set of linguistic terms to discuss a topic, then the language diffusion might be inflated for individuals. Future analysis should remove commonly used words in a group in order to identify unique terms that leaders produce. More generally, the way in which the language scores were averaged for each author diminishes some of the immediate language effects that could be associated with a message reply; future work should look more specifically at the language between linked messages.

Most importantly, Google Groups (and Usenet) represent a distinctive use of computer-mediated communication in which “participants, all of whom are involved voluntarily, often transform informal links into distinctive intentional sub-cultures” (Baym, 1995, p. 30). Even more so, Tepper (1997) argues that different discussion groups attract distinctive readers and posters from the larger heterogeneous Usenet, and adherence to group norms and standards leads to unique subcultures. These cultural practices may contribute to unique online interactions that may not always
be evident in other online communities. Therefore, future research should apply these models to other social media contexts.

In conclusion, this research provides a rich understanding of leadership and social influence in online communities by identifying a set of traits that increase an individual’s ability to influence the communication behaviors of others. At a much higher level, it reflects many of the advantages of mediated communication and virtual communities outlined by Katz, Rice, Acord, Dasgupta, and David (2004). Namely, online discussion groups allow users to build heterogeneous and often spontaneous communities, to increase social contacts and social capital, and to develop meaningful interpersonal bonds on a large scale. This study shows that online communication activity and the flow of information are strongly associated with social capital and reciprocal exchange, with intimate and affective interaction, and ultimately, through a sense of community that drives social interaction and social influence.

Acknowledgments

The author thanks his dissertation committee, which includes Noshir Contractor, Daniel Diermeier, Eszter Hargittai, and Darren Gergle, as well as Marc Smith, Robert Kraut, Joseph Jorgensen, Moira Burke, Yun Huang, and York Yao. This research was funded in part by a Donald H. and Caroline E. Ecroyd Fellowship.

References


【摘要】

本文旨在探讨网上舆论领袖，或那些对社区其他成员有影响的网民的沟通行为，他们或启发回复，或启发对话，或传播语言。本文的数据来自于两年间谷歌社区16个讨论小组的33,450个参与者的632,622条信息。本文利用自动文本分析、社会网络分析和层级线性模型来揭示网上舆论领袖的语言和社会行为。调查结果表明网上舆论领袖通过频繁的在线交流活动、可信度、网络中心性以及利用情感、自信和语言的多样性来影响他人。
Dimensions du leadership et de l'influence sociale dans les communautés en ligne

David Huffaker

Le but de cet article est d'examiner les comportements communicationnels des leaders en ligne ou de ceux et celles qui influencent d'autres membres de communautés en ligne en provoquant des réponses à des messages, en déclenchant des conversations et en propageant des styles linguistiques. L'étude s'appuie sur 632 622 messages de 33 450 participants à 16 groupes de discussion basés dans Google Groups au cours d'une période de deux ans. Elle utilise l'analyse de texte automatisée, l'analyse de réseaux sociaux et la modélisation linéaire hiérarchique pour révéler le langage et le comportement social des leaders en ligne. Les résultats montrent que les leaders en ligne influencent les autres par leur forte activité de communication, leur crédibilité et leur centralité au réseau, et par l'usage d'affect, d'affirmation de soi et de diversité linguistique dans leurs messages en ligne.
Dimensionen der Führung und des sozialen Einflusses in Online-Gemeinschaften

David Huffaker

온라인 커뮤니티들에서의 리더십차원들과 사회적 영향력

David Huffaker

요약

본 논문의 목적은 온라인 리더들의 커뮤니케이션 행위들, 또는 메시지 응답들을 이용하거나, 대화를 자극하거나, 그리고 언어를 전파함으로써 온라인 커뮤니티들의 다른 구성원들에게 영향력을 행사하려는 행위에 대한 연구이다. 이는 지난 2년동안 구글 집단들로부터 16개 모임에 걸친 33,450명의 참여자로부터 획득한 632,622메시지를 분석하여 나온 결과이다. 본 연구는 자동화된 텍스트 분석, 사회적 네트워크 분석, 그리고 계층적 선형모델링을 이용하였는바, 이는 온라인 리더들의 언어와 사회적 행위를 분석하기 위한 것이다. 연구 결과, 온라인 리더들은 높은 정도의 커뮤니케이션 행위들, 신뢰도, 네트워크 중심성을 통해 다른 사람들에게 영향을 주었으며, 이들은 온라인 메시지들에서 친화적이고, 단언적이고, 그리고 언어적인 다양성을 사용한 것을 발견하였다.
Las Dimensiones del Liderazgo y la Influencia Social en las Comunidades Online

David Huffaker
Media, Technology and Society, Northwestern University, Frances Searle Building, Room 2-148, 2240 Campus Drive, Evanston, IL 60208, USA

Resumen
El propósito de este artículo es examinar los comportamientos de comunicación de los líderes online, o aquellos que influencian a otros miembros de las comunidades online al desencadenar los mensajes de respuesta, generar conversación, y difundir lenguaje. Cuenta con 632.622 mensajes de 33.450 participantes a través de 16 grupos de discusión de los Grupos Google que se llevaron a cabo en un período de 2 años. Utiliza un análisis de texto automatizado, un análisis de la red social, y un modelo lineal jerárquico para descubrir el lenguaje y el comportamiento social de los líderes online. Los hallazgos muestran que los líderes online influencian a otros mediante la alta actividad de comunicación, la credibilidad, la centralidad de la red, y el uso de afectividad, la firmeza, y la diversidad lingüística en sus mensajes online.