NORTHWESTERN UNIVERSITY

Dimensions of Leadership and Social Influence in Online Communities

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Media, Technology and Society
School of Communication

By

David A. Huffaker

EVANSTON, ILLINOIS

December 2009

Abstract

Dimensions of Leadership and Social Influence in Online Communities

by David A. Huffaker

The purpose of this dissertation is to examine the communication behaviors of online leaders, or those who influence other members of online communities in terms of triggering replies, sparking conversations and diffusing language. It also examines the influence of group attributes on leadership such as size and connectedness. It relies on roughly 500,000 messages from 33,450 participants across sixteen discussion groups from GOOGLE GROUPS that took place over a two-year period. It utilizes automated text analysis, social network analysis and hierarchical linear modeling to uncover the language and social behaviors of online leaders. The findings suggest that online leaders influence others through high communication activity, credibility, reciprocal social network behaviors, and the use of affective, assertive and linguistic diversity in their online messages. Brokering, in which users connect to those who are not connected to each other, is not a significant predictor, suggesting that transparency and accessibility in online environments reduce the advantages of serving as a broker. In addition, group attributes such as size and network density encourage the emergence of leaders. However, participation equality and group turnover do not affect these behaviors, which emphasize the unique context of online communities, which often show power-law distributions of participation and high attrition rates. Taken together, the findings extend existing theories of social influence found in communication studies and social psychology, and increase our theoretical understanding of online leadership.

Acknowledgements

First, I must thank my excellent dissertation committee, Daniel Diermeier and Noshir Contractor, who co-chaired this work, as well as Eszter Hargittai and Darren Gergle. They all served as great mentors and guides throughout this process. Jennifer Lai (IBM Research) was also extremely helpful in shaping this work in its early stages.

I especially thank Marc Smith (formerly of Microsoft Research) and Robert Kraut (Carnegie Mellon University) for providing me with the data set. I also thank Moira Burke, Yun Huang and York Yao for assistance in transferring and setting up the data set.

I am indebted to Joseph Jorgensen for his programming help and guidance.

This work has been influenced by excellent feedback from folks at IBM Research (especially John C. Thomas and Rob Farrell), AT&T Research (especially Hal Purdy and Brian Amento), and the CSST Summer Research Institute (especially Steve Sawyer, Steven L. Johnson, Eric Gilbert and Derek Lackaff). Likewise, several of my graduate student peers provided keen insights along the way: David Miele; Paul Tepper; Jason Gallo; Tom Ksiazek; Gina Walejko; Sean Zehnder; and Mike Graziano.

Special thanks for Bob and Janice Horowitz for their editorial assistance.

I must also thank my loving partner, Emily. And thanks to my family, friends, and fellow graduate students for all their support and encouragement.

“The prophet has his disciples; the war lord his henchmen; the leader, generally, his followers.

There is no such thing as ‘appointment’ or ‘dismissal’, no career, no promotion.

There is only a ‘call’…” –Max Weber (1947, p. 360).

Dimensions of Leadership and Social Influence in

Online Communities

By David A. Huffaker

List of Figures

List of Tables

Chapter 1. Introduction

Online communities, in which large groups of internet users with common interests or activities communicate and share resources with one another (Preece, 2000), make up a considerable portion of internet use (Horrigan, 2001). They manifest as social networking services such as MYSPACE or FACEBOOK, content-sharing web sites such as YOUTUBE or FLICKR, online discussion groups such as GOOGLE GROUPS and all variety of message boards. Even blogs, inundated with comments from loyal readers or first-time viewers, represent a thriving 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.

Many scholars have hailed online communities as egalitarian spaces that facilitate powerful information exchange, provide emotional support or mobilize users into collective action (Baym, 2000; Rheingold, 2003; Wellman & Gulia, 1999). Online communities continue to spread in the work place (Sproull & Kiesler, 1991; Wallace, 2004) and educational settings (Pittinsky, 2003; Ryan, Scott, Freeman, & Patel, 2000), pointing toward their potential to promote innovation, collaboration and learning (Alavi, 1994; Finholt & Sproull, 1990; Kouzes, Myers, & Wulf, 1996). Understanding the complex social behaviors and group dynamics that take place in online communities and social media can provide considerable insight for social science.

For example, we know very little about who drives or leads these communities, or who serves as the most influential members of online groups. A deep understanding of the communication traits and social networks of online leaders not only informs previous theories of leadership in organizational settings, or provides a new lens for identifying social interaction or information diffusion, but it provides insights what makes online groups different than those in the physical world. Only recently have researchers begun to examine how leaders communicate in online groups (Cassell, Huffaker, Tversky, & Ferriman, 2006; Misiolek & Heckman, 2005; Yoo & Alavi, 2004), or how to track the flow of information within large-scale networks (Aral, Brynjolfsson, & Van Alstyne, 2007; Godes & Mayzlin, 2004; Gruhl, Guha, Liben-Nowell, & Tomkins, 2004). Very little, if any, work has been done that examines the combination of the two — that is, the relationship between leadership and diffusion in online communities.

Furthermore, while scholars have made many advances in understanding language and social interaction in computer-mediated environments (Crystal, 2001; Danet & Herring, 2007; Hancock, Curry, Goorha, & Woodworth, 2005; Niederhoffer & Pennebaker, 2002; Paolillo, 2001), as well as focusing on the communication and social networks found in online settings and their impact on behavior (Castells, 1996; Garton, Haythornthwaite, & Wellman, 1999; Rheingold, 2000; Turner, 2005; Wellman, Salaff, Dimitrova, Garton, Gulia, & Haythornthwaite, 1996), the two have never been combined to develop a theoretical understanding of online leadership.

Research Objectives

The purpose of this dissertation research is to investigate the communication behaviors that best represent online leaders, focusing on the language they use in their online messages, and the interaction patterns they engage in with other members of the group. Second, this dissertation explores the characteristics of the groups that leaders inhabit and their impact on social influence. It is motivated by two research questions:

1. What are the primary communication traits, including both linguistic characteristics and social interaction patterns, associated with online leaders?
2. To what extent do characteristics of the virtual groups moderate or enhance online leadership and social influence?

To answer these questions, I rely on a unique set of online behavioral data: a complete set of social interactions and online messages from a collection of GOOGLE GROUPS over a two-year period. This amounts to roughly 500,000 messages from over 33,000 participants. This data is special because it isn't obtrusive; these are the natural online behaviors of online users, free from moderation or the awareness that someone is watching from outside the community.

I also utilize a multi-method approach that encompasses techniques found in computational linguistics and social network analysis. I analyze the online messages of each participant using word frequency analysis to pinpoint psychosocial and cognitive dimensions of language. I identify the communication networks — who talked to whom — over the entire time period. I measure the extent to which participants spark long and multi-threaded dialogues, as well as how their word choices spread throughout the community. I utilize these approaches to develop a theoretical model of how leaders communicate and interact in online communities, and to measure their level of social influence. Finally, I use multi-level statistical evaluation to determine how group dynamics can influence the emergence of leaders and diffusion.

The purpose of my dissertation is to make descriptive and theoretical contributions to our understanding of the ways in which leaders communicate in online settings and the level of influence these communication traits have on other members of the community. Second, I aim to make a methodological contribution by demonstrating the utility of text analysis and social network analysis in studying online behavior, as well as the importance of considering group-level variables when examining individual behavior. Such quantitative measures offer a unique lens for studying complex social behavior. In effect, my dissertation fills an important void in computer-mediated communication. It provides a comprehensive framework for understanding leadership in large-scale online networks, and a set of analytical tools that can uncover behavioral phenomena that are typically overlooked.

I'll argue that online leaders engage in more communication activity than their counterparts, and that they also utilize more relationship-building and bonding behavior instead of merely broadcasting their opinions. Second, leaders demonstrate more credibility to their peers through tenure in the community and linguistic competence. Third, leaders demonstrate language that is rich with emotion, confidence and assertiveness, all of which increase their ability to influence others. Finally, I show that group attributes such as its size and connectedness improve social influence in online settings.

Significance of the Research

The results of my research contribute to several areas in social science. First, it enhances our understanding of leadership theory by identifying aspects of discourse and social structure that distinguish leaders in a group or organization, and communication theory by identifying mechanisms that facilitate diffusion in computer-mediated settings. Second, it demonstrates a comprehensive quantitative analytical approach to examine user behavior in large-scale complex networks. Third, it develops interdisciplinary bridges for future research by integrating research in communication studies, organizational science, social psychology, computational linguistics, and social network theory. This study also addresses several subsidiary issues:

- Developing a specific linguistic and structural model that can be used to identify leaders in large-scale networks. For example, it can be used to identify leaders within a social networking service such as FACEBOOK, or determine the most influential bloggers. Targeting leaders is important for community managers and advertising systems alike.
- Improving the design of online communities to facilitate user collaboration and communication; e.g., the building of recommendation systems that link users with similar linguistic traits or social interaction patterns, such as a

recommendation system that can match users of an online dating site, or identify users with similar product tastes, or locate experts within the community.

- Incorporating group dynamics into the study of individual behaviors. For example, comparing the complexity and turnover of various groups can show technology designers where to fine-tune an application, or when to splice a large group into smaller, more manageable pieces. Second, group-level variables can be included when comparing several communities from the same genre (e.g., FRIENDSTER vs. MYSPACE vs. FACEBOOK). This way, researchers can identify technical and sociocultural features that distinguish one application from another, possibly explaining popularity or success.

As online communities continue to expand on the internet, as well as within organizations or schools, understanding the complex social behaviors that comprise them provides insights for scholars, managers and technology designers alike. For social scientists, examining communication processes in online contexts helps with building theories around computer-mediated communication, which is still in its infancy. For managers, this work can serve to uncover employees who are actively contributing to the success of the organization and emerging as leaders, but are hidden within the hierarchy. For technology designers, analyzing social behavior allows for improving online applications and user experience.

Thesis Overview

The general approach of my dissertation is to identify leaders based on various aspects of social influence and then empirically test these assumptions in online discussion groups. Chapter 2 provides the conceptual framework for understanding leadership and how these theories translate into aspects of language use and social network patterns. It includes a review of empirical work in communication studies, social psychology, organizational science, linguistics, and sociology to provide an interdisciplinary background for understanding leadership in online and real-world contexts.

After presenting the models and hypotheses of what contributes to online leaders, I develop a series of measures that can be used to empirically measure these behaviors and characteristics. Chapter 3 begins by my description of the data set: a massive collection of online messages and interactions from sixteen discussion groups on GOOGLE GROUPS, which is the current manifestation of *Usenet*, one of the oldest internet communities still active today (Baym, 1995). Next, I explain the research design and variables of interest. I introduce text analysis and other techniques in computational linguistics that are used to measure human language, and also introduce social network analysis and several measures of centrality that are commonly used to identify the most popular or powerful members of a group and evaluate a person's social capital. I present several new metrics for examining group attributes in online settings.

In Chapter 4, I present a detailed analysis of each hypothesis, starting with descriptive data that illustrates the group dynamics and individual communication traits that are at play in this community. I rely on hierarchical linear modeling (HLM) throughout most of the analyses in order to measure both individual-level variables such as language use and social interaction, as well as group-level variables such as size and stability. HLM is especially useful here because it allows us to separate the individual-level effects from the dynamics of the group, a statistical approach that has been recommended for use in communication studies (Slater, Snyder, & Hayes, 2006).

Finally, I summarize the results of the data analysis and conclude the dissertation in Chapter 5. In it, I discuss the findings in light of previous computer-mediated communication (CMC) research. I also describe the theoretical, methodological and practical implications of this research. I finish the chapter by pointing out the limitations and future directions of this work.

As online communities continue to grow and remain popular meeting places for internet users, understanding human behavior in these group settings can provide considerable insight for social science. Overall, my research will serve as foundational work for communication scholars, computational linguists, and human-computer interaction researchers to build upon.

Chapter 2. Conceptual Framework

Understanding Leadership in Online Settings

Online leadership is a unique phenomenon. It does not fit neatly into any of Weber's (1947) archetypes of leadership. Online leaders do not represent legal authority, in which they are elected or appointed[1], nor do they represent traditional forms, in which they inherit a position of power (Weber, 1947). Likewise, online leaders are not necessarily "charismatic" or "…set apart from ordinary men and treated as endowed with supernatural, superhuman, or at least specifically exceptional powers or qualities" (Weber, 1947, p. 359). Rather, online leaders tend to emerge in otherwise leaderless groups, but they do make an impact on the behaviors or attitudes in the online spaces they inhabit.

Leadership is really about followers. Without other members allowing themselves to be influenced, important information could not be diffused within a group, nor would users mobilize around particular topics or ideas. And of course, no individual responsible for spreading ideas or information would exist. The leader-follower relationship, which has enjoyed a considerable amount of scholarship in leadership studies (Bass, 1990), remains dynamic: leaders can become followers and vice versa; and multiple leaders can arise and coexist within a group (Jago, 1982). This is exemplified in online settings, where people

[1] There are communities where some members serve as appointed moderators, making sure that all discussions follow a set of behavioral norms. However, these are exceptions to the rule; the majority of online groups and communities do not include appointed moderators.

share content within local and global communities at an unprecedented level. The span and speed at which news articles, entertaining videos, email scam warnings — essentially hot topics or opinions — travel, has everything to do with social influence and those who are influential.

Therefore, online leadership is best understood in terms of *influence*. As Hollander (1961) states: "…a leader denotes an individual with a status that permits him to exercise influence over certain other individuals" (p. 30). In a more recent work, Jago (1982) describes leadership as a process involving "…the use of noncoercive influence to direct and coordinate the activities of the members of an organized group toward the accomplishment of group objectives" (p. 315). Similarly, Turner (1991) argues that:

> "Leaders are persons or social roles who exert more influence in a group than others…they tend to suggest, direct, instruct and advise courses of action that the group actually follows. They play the most important role in directing the group's activism, maintaining its traditions and customs and ensuring that it reaches its goals" (p. 132).

The idea of leaders as influential members of a group or community translates nicely on the internet because some individuals emerge as the most important members and influence the communication behaviors and content of the rest of the group. Members of an online community focus on particular topics (e.g., a discussion group on cancer support, or a fan club group) or a particular purpose (e.g., connecting with old friends, keeping everyone abreast of the latest doings, sharing and rating news stories). In a topic-based community, influential members will steer the conversation by bringing up topics that the rest of the group finds interesting or controversial, inspiring conversation or debate. In a news-sharing and rating forum such as DIGG, influential members will contribute stories that result in high readership, and again, discussion. In both examples, influential members of an online community impact—and perhaps shape—public discourse. They do so by promoting discussion, and influencing the actual content of these discussions.

Even the most individual pursuits such as writing a blog or constantly updating one's status via social or instant messaging applications are only as successful as the followers and comments that the blog posts or feeds acquire. When we think of prominent bloggers, we think of those who have the largest following, whose posts are syndicated in other venues, or who generate thousands of comments. Advertisers seek popular online users because they represent *opinion leaders*, and marketers hope that ideas or products embraced by influential users will result in more adoption by the larger community.

Opinion Leaders

This concept of opinion leadership has a rich history in communication studies. Opinion leaders, which can be found in all age groups, ethnicities, and across gender and socioeconomic status, are often characterized as being highly interested and informed in their various domains of interest (Weimann, 1994). Whether it's spreading messages from the mass media (Katz & Lazersfeld, 1955) and diffusing innovation (Rogers, 1995), or acting as knowledgeable sources within small local networks (Weimann, 1994), opinion leaders are considered influential because they are able to gather information from social channels and then disperse that information or advice through such channels. (Weimann, 1994).

For example, opinion leaders within the political sphere tend to closely follow political issues in the media and then frequently discuss these issues with other people in various discussion forums (or even one's next-door neighbor) with the hopes of shaping the opinions of others (Weimann, 1994). They inform themselves and then broadcast their ideas through their communication networks. In scientific domains, opinion leaders are scholars who hold central positions in the community and shape scientific progress through a "…powerful 'invisible college' that dominates the adoption or rejection of new scientific models, ideas and methods" (Weimann, 1994, p. 205) [2]. Likewise, opinion leaders in health care domains (often health care professionals influencing patients or clients) either promote or block innovation, new technologies or practices (Weimann, 1994).

These domain examples are especially relevant because the spheres of influence outlined by communication scholars are represented all over the internet. In fact, the most popular online discussions are focused on politics, health, hobbies or science and technology issues (Horrigan, 2001). Searching for any topic will inevitably lead to a discussion forum or message board dedicated to it, and all major sources of information, from the NEW YORK TIMES to WEBMD have incorporated community discussion features to their site to encourage participation and information sharing.

---

[2] Scientific community is an excellent example of how followers create leaders. For example, scholars are often determined to be preeminent based on the number of citations they receive from other members of the community (Garfield, 1979).

Because opinion leaders are often the gatekeepers of information, and other members of the group tend to go to them for information or advice (Burt, 1999; Rogers, 1995), online users often keep influential individuals close. Users can follow influential members through social networking services, micro-blogging applications (e.g., TWITTER) and instant messaging (IM) systems. Users can also bookmark sites or blogs they enjoy, or subscribe to influential individuals via RSS feeds[3], which allows them to connect with content through email or mobile phone, and be able to comment.

Agenda Setting and Framing

Online leaders are able to influence others in the community in a couple of ways: First, they shape what people talk about by stimulating or facilitating discussion on a particular topic or phenomenon; and second, they shape *how* people talk about a topic or phenomenon. We can think of the first type of activity as agenda-setting (Dearing & Rogers, 1997). That is, leaders post messages that catalyze discussion or share stories or videos that generate many responses and feedback. These agendas appear in blogs, on message boards, in social networking services or social messaging applications. They can be topical (e.g., current news, a new movie, controversy) or of general interest (e.g., how-to / Do-it-Yourself communities, fan clubs, spam alerts) but in all cases, the topics must have an original author or source, and the potential for prolonged discussion. This is evident in the myriad

---

[3] Really Simple Syndication (RSS) allows user to capture stripped down text versions of blogs and other web applications, which can be reviewed on the web, in email or on a portable device.

of 'tag clouds'[4] that illustrate the most popular keywords or discussion topics existing in a given community. For instance, a review of TWITTER's search feature shows that conversational trends include 'AIG', 'social media' or 'iphone' — each of these trends has an original source.

The second conceptualization involves the ability to spread a concept or shape a topic. We can think of this type of activity as framing (Conger, 1991). For example, leaders might emphasize certain aspects of an issue in their claims and concepts, which get picked up and repeated by other members in the conversational thread, or redistributed to other venues (e.g., syndicating a blog post). The spread of a concept occurs when an individual begins talking about a new product or activity, introducing a new term that other members learn and re-use. This is also captured in Dawkin's (2006) notion of *memes*, or "tunes, ideas, catch-phrases, clothes fashions, ways of making pots or of building arches" and imitation, in which these ideas spread from "brain to brain" (p. 192).

There are many examples of this on the internet. First, there is an extensive lexicon created by internet users to represent technical terms (e.g., 'podcasting', 'blogging','MUD', 'FAQ), behavioral phenomena (e.g., 'netiquette, 'lurking', 'trolling', 'AFK') or emotional expressions or slang (e.g., 'LOL', 'IMHO', 'w00t') which have been adopted by the general

---

[4] Tag clouds are graphical representations of words and phrases found on web sites. Typically, they are represented as a list of words in alphabetical order, and frequently used words are differentiated by size or color. In effect, the most used words will appear largest in size. See http://en.wikipedia.org/wiki/Tag_cloud for examples.

internet population (Crystal, 2001). Second, users can shape topics by using a particular set of words to capture a concept or idea. They might introduce new terms that other members mirror, or direct conversation by using a specific set of key words used to capture a concept.

For example, in a political discussion, a leader might utilize a new term such as "republi-goon" which other members adopt and re-use in the conversation. They may use a specific example such as "the Wisconsin public school system" to discuss government spending, or use a phrase such as "toxic assets" to illustrate a larger fiscal issue, which other members adhere to when they continue the discussion. Sociolinguistics has studied this phenomenon in terms of "conceptual pacts", in which people engaged in a dialogue will come to an agreed-upon set of terms to represent mutual understanding (Brennan & Clark, 1996, p. 1482). This is illustrated in the co-development and discussion surrounding WIKIPEDIA content: as many users co-author the encyclopedia article, they must converge on a set of concepts or rhetoric to discuss the topic, and the editors rely on a discussion forum to manage this convergence.

What Online Leaders Are Not

There has been little empirical work studying online leaders. Most of this has focused on the 'babble hypothesis', which suggests that leaders emerge in virtual groups based on their frequency of communication or sheer volume of speech (Misiolek & Heckman, 2005; Sudweeks & Simoff, 2005; Yoo & Alavi, 2004). While these findings are reflected in some communication traits associated with opinion leaders (discussed in more detail below), critics argue that the communication must be relevant to the group, that quality is more important than quantity, and that "babblers are often squelched" by the group majority (Bass, 1990, p. 93).

There are special types of internet users that influence the community but do not serve as leaders. First, *trolls* are users that purposely post provocative, often emotionally-laden messages, to incite heated discussion (Donath, 1998). They may be able to spark conversation, but they aren't leaders. As Jago (1982) points out, "a person shouting 'Fire!' in a crowded theatre may indeed be influencing many lives [but…] would not be exhibiting leadership" (p. 316). Instead, the leader is the person who can calm the crowd and facilitate the evacuation (Jago, 1982). In other words, while trolls are in a sense 'babbling', they do not contribute relevant material to the group. And while trolls were a powerful disruptive force in early discussion forums, users began to regulate trolling through quick identification, ignoring or blocking messages that look like bait (Herring, Job-Sluder, Scheckler, & Barab, 2002).

Second, *spammers* are posters who contribute a high frequency of communication though advertisements, binary posts (i.e., images) or other nonsensical content. They flood the discussion group with off-topic content (Hoffmann, 2007), but typically receive no response (Turner, Smith, Fisher, & Welser, 2005). However, they may influence users with their content, or provoke unwritten response (i.e., annoyance), but they do not represent leadership. The point here is that trolls and spammers reinforce the idea that leadership cannot be represented by communication frequency or contribution alone; an examination of their impact on other users provides a deeper understanding of their role in the communities.

Summary

In this dissertation, I define online leadership in terms of its influence over others. I conceptualize this in terms of both agenda-setting and framing. That is, online leaders influence *what* people talk about. They create hot topics, such as sharing important news stories that other users spread or comment on, or offer their own comments or critiques that create a long discussion. Second, online leaders influence *how* people talk about it. They frame online messages in particular ways, emphasize certain aspects of a topic or phenomenon, and use specific types of language that other users pick up and spread down the conversation thread.

Agenda-setting and framing represent social influence in terms of "the process by which individuals make real changes to their feelings and behaviors as a result of interacting with others who are perceived to be similar, desirable or expert" (Rashotte, 2007, p. 4426), but these types of influence are by no means independent. Instead, they are most likely interrelated; leaders can both spark conversation and cause people to mirror or re-use content. But it's also possible that a leader can initiate a long discussion that does not reflect their original language; and a leader could create a short series of replies that completely embrace their language use.

While there has been little empirical work regarding online leadership, the intersection of communication studies, social psychology and management science provides a strong justification for the conceptualization of online leaders as influential members of a group or community. In the next section, I'll examine factors that positively impact social influence in order to understand the mechanisms that online leaders might employ to get their point across.

Factors that Increase Influence

The factors that increase influence for online leaders can be characterized in several ways. These include the attributes routinely associated with the individual contributor (often referred to as the '*source*' of messages) such as communication patterns and credibility. They also include characteristics of the *content* of online messages such as language clarity, power and intensity. In addition, attributes of the group and the recipient of messages can also impact influence. While theories of social influence (and persuasion in particular) often deal with compliance (i.e., a direct request) or attitudinal change (Cialdini, 1994; O'Keefe, 2002) in special settings such as health communication or marketing, they still serve as an important theoretical foundation for identifying features that would increase or hamper the spread of information in an online community.

Sociability and Network Centrality

Research on opinion leadership in communication studies notes several traits or behaviors that promote influence. One trait is simply *sociability* or gregariousness. As Weimann (1994) points out, "…opinion leaders in every domain are socially very active; they come into contact with many people…they speak at meetings, participate in discussions, and take part in many social events. They are well integrated into many social networks and have many friends and acquaintances" (p. 78). This coincides with the babble hypothesis by showing that leaders communicate frequently; however, further work shows that it's not just about talking, but to whom you are talking.

This focus on social networks shows that opinion leaders are often the most central figures or "sociometric stars" in their corresponding networks (Weimann, 1994, p. 81). Opinion leaders are not only more active in their communication frequency, but also their communication reach. They rely on far more personal communication than non-leaders to find information or keep up-to-date on new trends and innovations, and then to diffuse their ideas or opinions (Weimann, 1994). Regardless of the domain (e.g., fashion, hobbies, politics or health) opinion leaders are influential through high levels of social activity, leaving centrality as an important research area in the understanding of influence (Weimann, 1994).

The Credibility of the Source

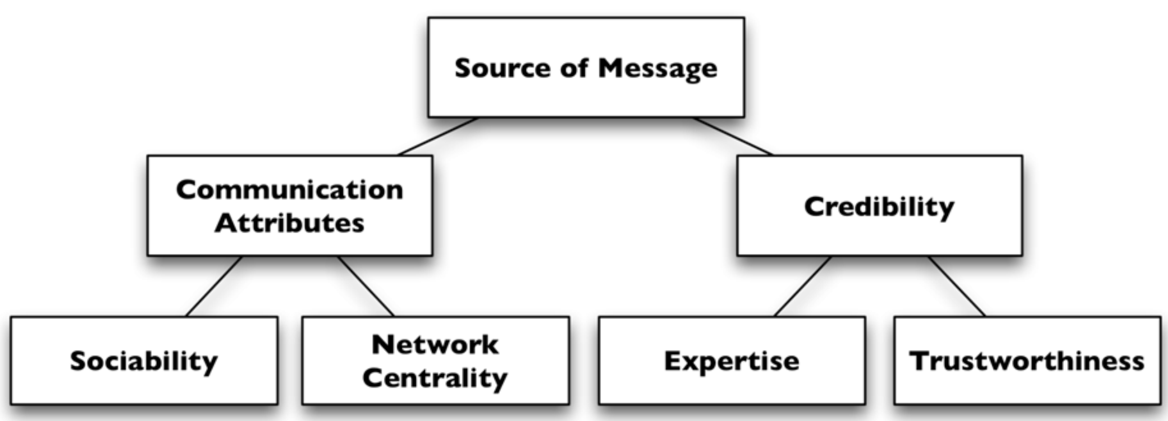
Second, scholars have noted that the *credibility* of the individual (who is often referred to as the 'source' of influential messages) is very important in increasing influence (O'Keefe, 2002; Pornpitakpan, 2004). This includes both the *expertise* of the source and their *trustworthiness* (O'Keefe, 2002). Expertise represents "competence, expertness, authoritativeness or qualification" and is often understood in terms of experience, skill and intelligence (O'Keefe, 2002, p. 182). While leaders usually show expertise in only one domain or topic area, this is a central reason why people go to them for information or respect their opinions (Weimann, 1994). Trustworthiness represents "character, safety or personal integrity" and refers to honesty, open-mindedness, fairness and the inclination to tell the truth (O'Keefe, 2002, p. 183).

Trustworthiness is also related to group membership, or the length of time spent within the group. Hollander (1961) argues that individuals can only attain leadership status when they are in a group long enough for others to recognize their contribution to the group goals or purpose, stating "it is unlikely that just anyone in the group could achieve leadership by a suggestion for change at an early stage of membership…this is the dilemma of the neophyte [who] is most restricted of all" (p. 38). This resonates with social impact theory, which posits that influence is a function of the number of individuals that constitute a source, how close they are to other members and their salience, status or power (Latane, 1981). Being a recognized or dedicated member of a group increases one's immediacy and strength, and studies show that influentials thrive on social recognition and credibility (Weimann, 1994).

Both expertise and trustworthiness are reflected in online communication. For example, studies show that users rely on experts to answer questions, get advice or find useful information (McLure Wasko & Faraj, 2000). In order for individuals to appear credible, their messages must appear informed and knowledgeable, perhaps even well written. In addition, these online sources must also be considered trustworthy; there should be no indication that the author is speaking dubiously or facetiously (again, characteristics of a troll). Trustworthiness is enhanced when the source appears embedded in the group and dedicated to its success. Many online communities reinforce this aspect by providing additional information such as the length of membership in the community (i.e., 'Member since 1996') or user status (i.e., 'Super User'), both of which distinguish an individual with an exceptional tenure in the community. Even when these attributes aren't directly available, users can usually determine how long someone has been around based on the time-stamps associated with each post.

Figure 2-1 illustrates the most popular attributes of social influence associated with the source of a message. These are comprised of communication attributes such as sociability and network centrality. They also involve credibility, which includes expertise, competence,knowledge and also trustworthiness. I posit that each of these factors remains important for online leaders.

Figure 2-1. Source factors that increase social influence.



Clarity, Power and Intensity in Message Content

In addition to characteristics of the authors of persuasive messages, scholars have studied the outcomes of individual messages including message structure, message content and “sequential-request” (a.k.a. compliance) strategies (O'Keefe, 2002, p. 215). Message structural features include the order of the arguments (which seem inconsequential according to empirical finding), whether the point or recommendation in the message is explicit or implicit (findings suggest explicit is more persuasive), or whether the point or recommendation is general or specific (findings suggest that specific descriptions are better) (O'Keefe, 2002).

Research on the actual lexical items used in the content of messages has focused primarily on message clarity, powerful and powerless language, and language intensity (Ng & Bradac, 1993). Message clarity, which includes the complexity of the message and the comprehension difficulty (Hosman, 2002), has been studied in terms of *lexical diversity* and *fluency*. Lexical diversity, which represents “vocabulary richness or vocabulary range” (Hosman, 2002, p. 374) represents complexity in terms of language. As Bradac, Konsky & 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, articulation difficulties” etc. (O'Keefe, 2002, p. 185). In other words, poor language ability or linguistic diversity negatively impacts credibility and influence.

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). However, some studies suggest there is no difference in power style, although "stronger" arguments are more persuasive than "weak" ones (Gibbons, Busch, & Bradac, 1991, p. 115).

A third area of research on the effects of message content on influence involves language intensity (Ng & Bradac, 1993). Hamilton & Hunter (1998) define language intensity as a "…stylistic feature of language that is conveyed through the properties of emotionality…" (p. 100). Others have defined it as "…language indicating degree and direction of distance from neutrality" (Burgoon, Jones, & Stewart, 1975, p. 241). In other words, language intensity is about salience and how well it 'stands out'.

Language intensity is often understood in terms of *emotion*. According to Hamilton and Hunter (1998), "…emotional intensity is the degree of affect reflected in the source's language, 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). Affect impacts "…the quality and originality of persuasive messages" (Forgas, 2006, p. 284) and as message recipients process message content, positive and negative framing can influence their decision to adopt or believe (Block & Keller, 1995; Maheswaran & Meyers-Levy, 1990). Thus positive or negative framing grabs the attention of a recipient, which allows the message to be scrutinized since it is more salient (Smith & Petty, 1996)[5]. The extreme example of this are fear appears, in which the message claims that if a person doesn't accept the recommendation a terrible fate awaits them. Fear appeals are generally considered to be a successful persuasive strategy.(O'Keefe, 2002).

O'Keefe (2002) also notes an active research area regarding "one/two way" messages, which address or ignore counterarguments (the empirical findings suggest that there is no difference unless the message directly refutes the opposing argument) (p. 219). Message content approaches have also included the frequency that recipients are exposed to messages, which can increase influence (Cacioppo & Petty, 1982; Cacioppo & Petty, 1989).

In sum, the study of message content, which often affects the perception and credibility of the source (Cacioppo & Petty, 1982), has focused on the clarity, intensity and power of the messages. Research shows that messages that are easily understood, or contain enough lexical richness to be interesting, contain enough intensity to grab attention or make an emotional appeal, or appear powerful enough to affect impressions, are typically more influential than their counterparts. Hamilton & Hunter (1998) argue that this

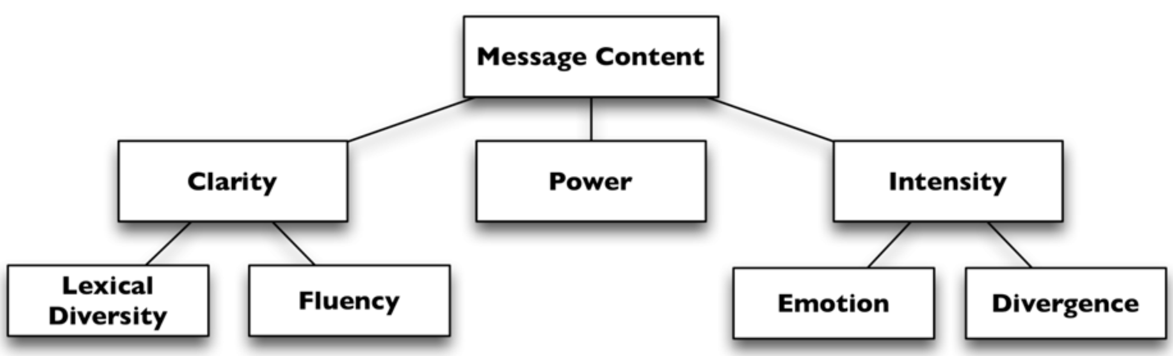
---

[5] Research also shows that the mood or disposition of the receiver can impact the processing of the persuasive message (see Bless, Bohner, Schwarz, & Strack, 1990).

is because language clarity, intensity or diversity impacts the perceived "dynamism of [the] presentation", which in turn impacts perceptions on source expertise and trustworthiness, and thus the attitude toward the source (p. 106). Of course, all of this is mediated by the knowledge of the receiver when evaluating a message—high knowledge results in a more critical review (Eagly & Chaiken, 1993).

Figure 2-2 illustrates the common properties associated with message content that increase social influence. This includes message clarity, which is comprised of lexical diversity (which demonstrates richness and complexity) and fluency (which can be thought of as error-free or high-quality). Powerful language has also been associated with social influence. Finally, message intensity, whether occurring from emotional or affective tones, or from the distance between the message and the recipient's attitudes or beliefs, also affects social influence.

Figure 2-2. Message content factors that increase social influence.



The 'Weapons of Influence'

In addition to message-centered approaches, scholars have outlined a broader set of strategies that foster influence. In particular, Cialdini (1994) identifies a handful of specific strategies that serve as "weapons of influence" (p. 1) that increase compliance (i.e., a response to a direct request). The first is *reciprocation*, in which people tend to "repay, in kind, what another person has provided us" (Cialdini, 1994, p. 17). That is, when people receive a favor or a gift, they will feel indebted and obligated to repay it. The exchange does not have to be fair; sometimes persuaders will ask for a small favor first, followed by a larger one, or vice versa, ask for a large favor first, get a rejection and then ask for a small favor[6].

The second compliance strategy, referred to as *commitment/consistency*, suggests that if people commit in writing or even orally, they tend to honor this agreement due to a pressure to appear consistent with previous actions or decisions (Cialdini, 1994). The third principle, called *social proof*, states that "…the greater the number of people who find an idea correct, the more the idea will be correct" (p. 128). This works in a couple of ways: when people are uncertain, they rely on the behaviors of others to guide them; and people are more likely to follow the lead of individuals who seem similar to them.

Liking, in which people "…prefer to say yes to the requests of those they like" also works with complete strangers (p. 167) because people like those who are similar to them in

---

[6] This is referred to as the "Foot-in-Door (FIT)" or "Door-in-Face (DIF)" strategies, respectively (see O'Keefe, 2002, pp. 230-233).

terms of personality, background or opinions. Liking also involves familiarity and repeated exposure (in which liking increases), or conditioning and association, in which we pair positive or negative connotations with people delivering information. The *authority* principle states that people tend to obey authority figures, even if they are asked to do unappealing things. Even the symbol of authority such as a title (e.g., 'Dr.') or clothes (e.g., police uniform) can increase compliance. The *scarcity* principle holds that "…opportunities [that]… seem more valuable to us when their availability is limited" (p. 238), which includes deadlines ('time is limited') or numbers of items ('only three left'). The feeling of scarcity increases our desire for the product, and increases competition among many people vying for the same scarce resources (Cialdini, 1994).

In sum, research in persuasion and social influence usually focuses on behaviors in which people or institutions are purposely trying to persuade another person. It isn't clear to what extent online users expressly attempt to persuade others, although there are many examples where persuasion can take place. For example, in a heated political discussion, the goal may be to engage in debate and enjoy the battle, as well as to convince others of an advocated position. Posting information such as news articles, entertaining content, new products or places, etc., may be about information sharing, but it's also about turning people onto new ideas or objects. Rating products or movies isn't necessarily about pitching them, but it does serve to warn others of 'bad' options.

Many of Cialdini's principles can be understood—at least broadly—in online communities. For example, it seems possible that reciprocation occurs (when a user comments to a post, the poster might respond in kind), but we can't be sure. Prominent bloggers may receive hundreds of comments in one day, and never return the favor. Second, *commitment/consistency* seems like a reason that people will argue about a topic, or 'stick to their guns' throughout the entirety of the discussion. However, this might seem contraindicated to spreading influence and deserves more investigation.

Social proof is clearly evident in the speed and depth at which the latest social media technology is embraced. For example, one of the earliest social networking sites, FRIENDSTER, was all but abandoned and never even came close to the audience of MYSPACE or FACEBOOK. While it shared many of the same features, it never reached a critical mass. In fact, we see examples of critical mass all over the internet, e.g., the spread of popular YOUTUBE videos or other entertaining content or the actual choice of web applications. These network effects, in which "…products or services become more valuable as more customers use them…" is the internet's trademark (Porter, 2001).

Liking is especially interesting because only verbal cues and ideas are available to depict similarity. However, we might surmise that repeated interaction and familiarity with other users would increase this liking. Similarly, *authority* may manifest through expertise or longevity in the community, in which other users consider the oldest, most active, or most knowledgeable members as authorities who set the behavioral norms of the group.

Finally, *scarcity* does not seem as relevant online because everyone has open access to information and interaction and there are unlimited resources. However, one might argue that the recency of a current topic (i.e., a recent news story) has a fleeting characteristic. That is, users have a short timeline to spread or comment on a topic before it becomes dated or irrelevant.

Social Influence in Online Settings

Generally, research shows that when people use the internet to seek advice, they are influenced by source credibility, personalization (i.e., the extent to which the information was tailored to a user) and familiarity with the source (Briggs, Burford, De Angeli, & Lynch, 2002). Although this line of research focuses on users interacting with web sites and not other online users, these factors reflect some of the issues previously outlined.

Guadagno & Cialdini (2005) argue that there is little difference in the ability to persuade others between face-to-face (FTF) and computer-mediated communication (CMC), but find *authority* and *commitment/consistency* as the only principles of influence that have been empirically studied. Godes & Mayzlin (2004) argue that "…people make offline decisions based on online information, [and] that online conversations may be a proxy for offline conversations" (p.558). They show that information disperses through word-of-mouth networks in online communities (Godes & Mayzlin, 2004), a phenomenon that several other researchers have demonstrated (Gruhl, et al., 2004; Matsumura, Yamamoto, & Tomozawa, 2008; Valente & Davis, 1999). Bickart and Schindler (2001) even show that discussion forums have more influence than marketing-generated online messages for users looking or product information. Adkins & Brashers (1995) show that certain linguistic strategies can foster influence in computer-mediated communication contexts.

Very few studies have focused on leaders in particular. Some studies show that online leaders focused on products, started longer threads, introduced product information or created topics that inspired debate in discussion groups (Matsumura, Ohsawa, & Ishizuka, 2002b). Studies also show that the most influential bloggers tend to actively read and integrate the blog posts of other community members into their own contributions (Matsumura, et al., 2008). Even the email networks of real-world organizations reflect a hierarchical structure in which managers tend to be more influential (Matsumura & Sasaki, 2007). However, these studies do not disentangle the micro-behaviors that impact influence on other members of the community.

Other studies have focused on group dynamics. For example, studies show that a strong identification with the group-at-large increases social influence in CMC, and more so when all members are anonymous (Postmes, Spears, Sakhel, & de Groot, 2001). Along with being a part of a group for an extended period of time, being connected with many members of the group should also be taken into account. As Putnam (2000) claims, "…a well-connected individual in a poorly connected society is not as productive as a well-connected individual in a well-connected society. And even a poorly connected individual may derive some spillover benefits from living in a well-connected society" (p. 20). In other words, the social structure of a group can affect the influence of central individuals (Mizruchi & Potts, 1998), This includes size, complexity and stability of the group. And although this has not been well studied in CMC contexts, it should be.

Another approach to group attributes includes participation equality and dominance. Some studies show that CMC encourages the contribution of minority opinions, in which an individual or small group of individuals voices an opinion that challenges the mainstream opinion (McLeod, Baron, Marti, & Yoon, 1997). However, CMC may not be as egalitarian or equalizing as once heralded (Rheingold, 2000). Weisband, Schneider & Connolly (1995) find that high status members of an online group tend to dominate the discussion and influence others just as in offline settings, or make disproportionate contributions to group discussions (Weisband, 1994). However, much of this work has focused on small groups.

Summary

In conclusion, the factors that increase social influence include sociability, source credibility, several attributes of message content, group attributes and the recipients of messages. Online leaders tend to be central in their social networks, but their perceived competence (i.e., expertise, knowledge, authority) and trustworthiness by other members of the group has an important impact. Their use of language can make their messages more credible, interesting, or salient, and message clarity, intensity and power appear to better engage recipients. Aspects of the group, such as majority opinion, participation equality and tenure can influence the recipient (i.e., social proof), make way for a leader or improve his credibility. Finally, the knowledge and commitment of the recipients affect their interpretation of messages and their willingness to adopt a new idea.

Operationalizing Social Influence in Online Settings

I have argued that online leadership is best understood in terms of social influence. I have also introduced a collection of factors that increase social influence, including those attributable to both the content of a message and the author of that message. And finally, I have shown that social influence and diffusion occur in online settings.

Sociability and Trustworthiness in Communication Networks

The positions that individuals occupy in their networks, and the length of time they've spent in these networks improve their level of influence, which has commonly been labeled sociability and trustworthiness (or credibility). These factors can also be understood in terms of *social capital*, in which the social structures that actors inhabit affect their behavior or outcomes (Coleman, 1988). As Burt (2000) describes, "…individuals who do better, are somehow better connected…" (p. 349). In other words, the position that an individual inhabits in a social network often result in more benefits or resources (Burt, 2000). Studies show that at least in organizational settings, centrality relates to status, reputation and the level of social influence that leaders wield (Brass, 1984; Mehra, Dixon, Brass, & Robertson, 2006). Individuals with central positions in their networks tend to have more access to, and more control over, information and other resources (Balkundi & Kilduff, 2005).

Putnam (2000) highlights a difference between bonding and bridging social capital. Bonding social capital is "exclusive…inward looking…homogeneous" while bridging social capital is "inclusive…outward looking…" (Putnam, 2000, pp. 22-23). Putnam (2000) uses the example of a country club as bonding and the civil rights movement as bridging, and each has its own strengths (i.e., bonding creates solidarity and reinforces self identity, while bridging is good for increasing diversity and information diffusion). Bonding social capital is related to reciprocity, which can refer to a specific *quid pro quo* or a more generalized assumption that if a person provides a favor he will receive a return from someone else in the network or community (Putnam, 2000). Behaviors that endorse reciprocity such as symmetry, mutuality or bi-directionality are common in communication networks (Monge & Contractor, 2003).

Social capital is also related to gate-keeping (Gould & Fernandez, 1989). Leaders with central positions in their networks tend to have more access to, and more control over, information and other resources (Balkundi & Kilduff, 2005). Scholars have noted that acting as a bridge to otherwise weakly connected sub-groups provides a distinct competitive advantage in terms of information regulation and dissemination (Burt, 2000), and can be used to increase the collective performance of the group (Cummings & Cross, 2003). In organizational settings, having this access and control can be attributable to better individual performance, promotion opportunities, sense of belonging, lack of turnover, amount of power and ability to innovate (Sparrowe, Liden, Wayne, & Kraimer, 2001).

Network centrality is not actually concerned with whether the ties are incoming or outgoing, just that some individuals or actors are more visible or prominent (Wasserman & Faust, 1994). While scholars have pointed out the difficulty in assessing centrality in a network (Degenne & Forse, 1999), Freeman (1979) outlined three major approaches to measuring centrality, which includes both the sum of incoming and outgoing links between actors, and also *betweenness* centrality, which measures the shortest path, or geodesic, between non-adjacent members of a network. Betweenness centrality suggests that members on the geodesic might have some control over the interactions between the non-adjacent members, what Wasserman & Faust (1994) refer to as “actors in the middle” (p. 188).

For Freeman, members with high betweenness centrality have more control over, and better coordination of, the general communication and information flow of the group (Degenne & Forse, 1999). Burt (1992) extended this concept with the notion of ‘structural holes’, in which members can take advantage of the weakest ties, or non-adjacent nodes, by brokering or controlling information between them. Within networks, most information circulates within groups rather than between them, and individuals who connect the gaps between two groups by engaging the weaker connections which create the structural holes, have a strategic competitive advantage, especially in organizational settings (Burt, 2000). Therefore, leaders may take advantage of structural holes to increase power, or they may prefer to connect members around them to increase the collective performance of the group (Cummings & Cross, 2003). In voluntary online communities such as discussion groups, I would expect leaders to focus more on connecting members, resulting in a high betweenness centrality or brokering ability.

Message Content and Natural Language Use

Scholars have examined how language is used to represent aspects of identity or personality (Cameron, 1998; Eckert, 1989), negotiate social relationships (Goffman, 1990; Schegloff, 1998) or represent power (Bourdieu, 1991; Foucault, 1978). It is no surprise that language is closely tied to how effective leaders are in facilitating group processes or outcomes, as well as how they are perceived by others in the group or community (Bass, 1990). Discourse analysis, which refers to the study of written or spoken language (Brown & Yule, 1983), allows us to measure the ways in which people use language to represent cognitive constructs, cultural practices or personality traits (Pennebaker, Mehl, & Niederhoffer, 2003; Sapir, 1949; Wittgenstein, 1958).

Recent computational approaches offer a new lens for understanding natural human language. Computational linguistics or natural language processing (NLP), for example, combine statistical evaluation, data mining and clustering approaches to determine patterns and rules of language (Jurafsky & Martin, 2000; Manning & Shutze, 2002). Many of these techniques rely on concordance, in which all words in a text or dialogue are indexed along with their specific location (Biber, Conrad, & Reppen, 1998). This allows scholars to analyze word frequencies, identify important keywords, compare different usages of words, find phrases and index word lists (Biber, Conrad, & Reppen, 1998).

Word frequencies can be also be used to determine the most important words in a text (Corman, Kuhn, McPhee, & Dooley, 2002), compare and classify words between multiple texts (Kilgarriff, 2001), and attach semantic meanings to the list of words (Landauer, Foltz, & Laham, 1998; Miller, Leacock, Tengi, & Bunker, 1993). Similarly, word frequencies can be compared to pre-defined dictionaries representing linguistic or semantic forms. For example, software has been developed to count word frequencies and link them to rudimentary linguistic features, as well as to social and psychological dimensions (Hart, 2000; Pennebaker, Booth, & Francis, 2006).

Talkativeness

Extroversion involves being gregarious and talkative (Eysenck, 1967). Research demonstrates that interactions between leaders and other members of a group or organization often involve high levels of extroversion (Bass, 1990). Leadership has been associated with a greater amount of communication activity (Hollander, 1986), and the volume or frequency of communication to other members is an explicit measure of this activity. Sorrentino & Boutillier (1975) find that perceptions of leadership were strongly influenced by the quantity of verbal interaction. More recently, Riggio, Riggio, Salinas, & Cole (2003) find that leaders who spoke the most were elected leaders of a group, even if their social skills were not better than their peers.

Again, this has been dubbed the Babble Hypothesis and suggests that sheer quantity of communication should make a difference. In computer-mediated communication settings, talkativeness equates to the quantity of written words in online messages, or the average message length produced by participants.

Affect

Several studies link affect, which involves one's ability to perceive, understand, evaluate and use emotion to interact with others, to leadership (George, 2000; Zaccaro, Kemp, & Bader, 2004). This is not only because leaders tend to exhibit higher optimism and positivity in their communication (Hart, 1984, 1987), but also because the emotional displays of leaders have a strong influence on the perceptions, moods and performance of those around them (Brief & Weiss, 2002; Humphrey, 2002). For example, Newcombe & Ashkanasy (2002) find that group perceptions of the quality of a leader are correlated with the amount of positive affect or emotion that the leader emotes. Conversely, Lewis (2000) finds that negative affect had a negative impact on views of leadership effectiveness when compared to neutral displays of emotion. Bono & Illies (2006) find that leaders' positive emotions are linked with corresponding moods of the group. Pennebaker & Francis (1996) demonstrate that cognitive and emotional processes manifest in lexical choices. For example, a word such as "love" represents a positive emotional valence, while a word such as "hate" represents a negative valence.

Assertiveness

Assertiveness is often an encouraged quality in leadership because it involves expressing one's thoughts or feelings in a clear, direct manner, which can improve group performance, prevent mistakes or confront problems (Jentsch & Smith-Jentsch, 2001). In terms of language use, assertiveness has been conceptualized as a form of powerful language, in which speakers utilize certainty in their speech, avoiding more powerless forms of language such as hedge phrases, tag questions or inhibited language (Holtgraves & Lasky, 1999).

Research has demonstrated that leaders do use more powerful language. For example, Hart (1984) argues that political and religious leaders, corporate executives and social activists all exhibit higher frequencies of assertiveness, certainty and resoluteness in their dialogues with others. More recently, Hart and Childers (2004) demonstrated how verbal certainty has been frequently used in the last thirty years in presidential communication. Therefore, I'll argue that verbal certainty represents assertiveness, self-confidence or resoluteness for leadership.

Linguistic Diversity

Communication ability can be measured in terms of speech or writing quality (Sternberg, 1999), or in the complexity and diversity of the language used (Bradac, Bowers, & Courtright, 1979). As I'll describe in more detail later, vocabulary richness or diversity is commonly used as a measure of verbal ability or facility and corresponds with the message complexity and fluency mentioned in theories of social influence.

Final Word

Even though the study of social influence has enjoyed a long history in social psychology and communication studies, little work has focused on influence in large networks, or more specifically online communities. Certainly, social network scholars have identified several models of how information should flow in networks, whether a function of simple exposure (i.e., contagion models), or a threshold of adoption after a number of other members of the group do so (i.e., threshold or cascade models) (Bikhchandani, Hirshleifer, & Welch, 1992; Granovetter, 1978; Monge & Contractor, 2003); however, they have not uncovered the user behaviors that might increase diffusion. Watts & Dodd (2007) argue that social influence is far more complex in large networks because actors influence each other in a variety of ways, not just a two-way flow from opinion leader to follower. Given this challenge, the study of online communities offers a lens for understanding these complex social interactions,

While the study of leadership and social influence has received little work in online contexts, the research mentioned here serves as a conceptual framework for understanding this phenomenon. Leaders are best defined as those participants who are influential within an online group. Their sociability, credibility and the content of their messages moderate the level of influence. Their tenure within—and contribution to—a group enhances the leaders' trustworthiness and credibility. By using messages that exude clarity and complexity, power and confidence, intensity and emotional valence, leaders are able to capture the attention of recipients. Of course, the knowledge of recipients and the dynamics of the groups they inhabit, also moderate social influence.

Summary of Hypotheses

In this dissertation, online leaders are defined in terms of agenda-setting or framing. More specifically, I operationalize online leadership in three ways: (1) by their ability to trigger a reply when they post an online message; (2) by their ability to create a long conversation when they post or reply to an online message; and (3) by their ability to diffuse the actual language they use in their posts and replies. For each of the hypotheses listed below, I will test all three measures of online leadership.

Sociability and the Posting Behavior of Leaders

H1a. Leaders are more likely to post messages to the community.

H1b. Leaders are more likely to reply to the messages of others.

H1c. Leaders are more likely to have a longer tenure in their online groups.

Centrality and the Social Networks of Leaders

H2a. Leaders are more likely to be expansive in their connection with others.

H2b. Leaders are more likely to be reciprocal in their connection with others.

H2c. Leaders are more likely to serve as brokers in their connection with others.

The Language of Leaders

H3a. Leaders are more likely to demonstrate higher frequencies of talkativeness in their online messages.

H3b. Leaders are more likely to demonstrate higher frequencies of affect in their online messages.

H3c. Leaders are more likely to demonstrate higher frequencies of assertiveness in their online messages.

H3d. Leaders are more likely to demonstrate higher frequencies of linguistic diversity in their online messages.

The Effect of Group Attributes on Leadership

H4a. Group size will be positively associated with leadership.

H4b. Group participation equality will be positively associated with leadership.

H4c. Group density will be positively associated with leadership.

H5d. Group turnover will be negatively associated with leadership.

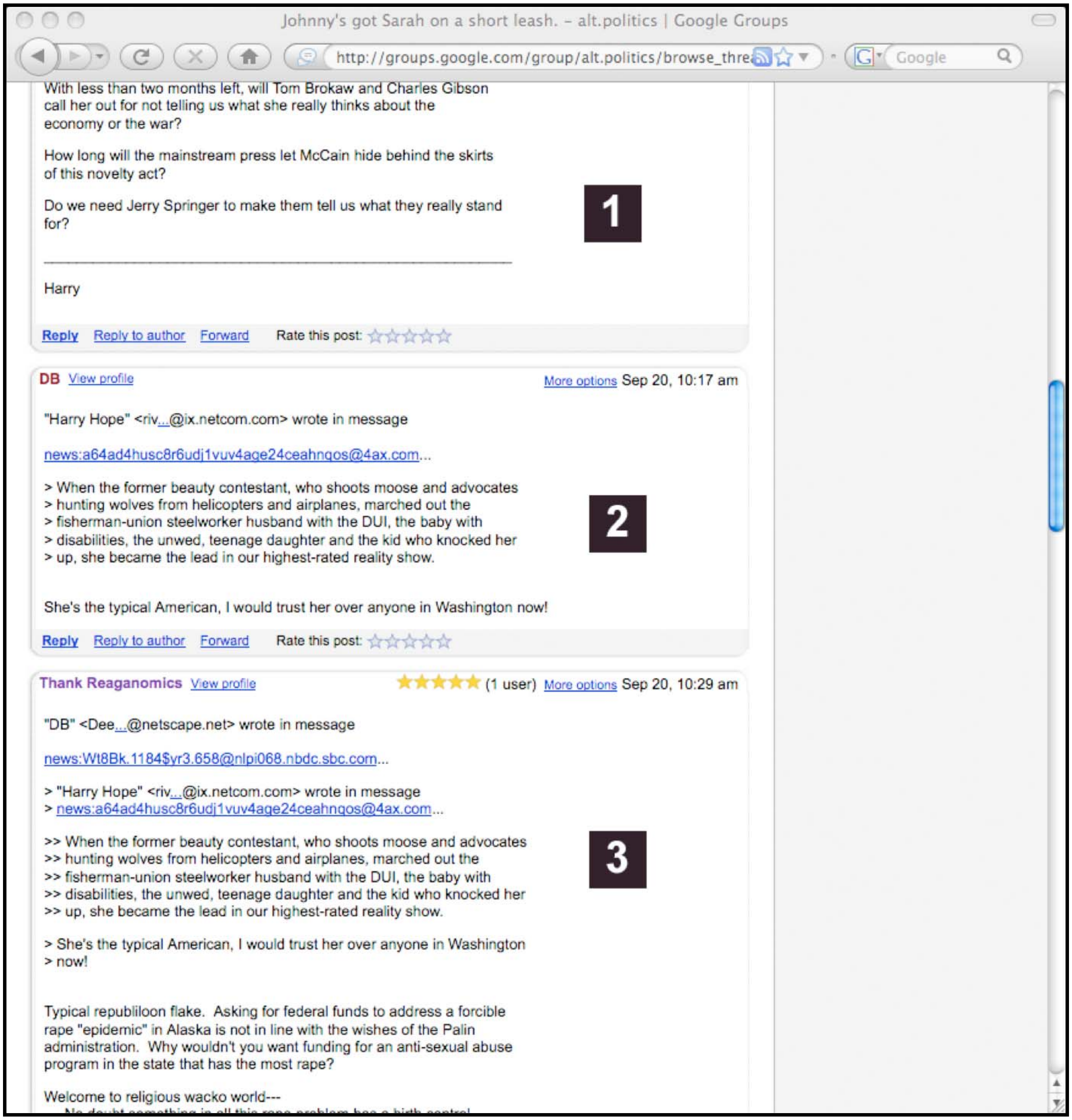
Chapter 3. Methodology

Sample

The sample in this study consists of 33,540 users who contributed 632,622 messages to a series of discussion groups found on GOOGLE GROUPS between 2003 and 2005. GOOGLE GROUPS is the current manifestation of *Usenet*, a web application that allows users to post messages to various topic groups and reply to the posts of other users, forming a conversational thread between two or more authors. *Figure 3-1* demonstrates a typical conversational thread found in GOOGLE GROUPS. The boxes show how three messages in the thread are formatted.

The message content, author name, date and time information are associated with each message. Note that several features such as the ability to post a user profile, or rate the postings (i.e., the "stars" in *Figure 3-1*) were not implemented until 2006. Therefore, they are not included in this study. The messages are ordered chronologically under a particular thread, identified by a subject. Users have the ability to forward a message, reply to the author of a message, or reply to the group itself, which forms a new parent thread. The messages often contain quoted material from a previous message to further distinguish which part of a message or thread the author is referring.

Figure 3-1. Example of a discussion group thread.



Note: The boxes marked "1", "2" and "3" represent three separate messages, ordered by date. They include a subject line (i.e., the colored text), the message and possibly some quoted text from a previous message. Users can reply or forward messages.

Netscan, a project at *Microsoft Research* (Smith, 2007), captured all social interactions that took place on *Usenet* over several years, focusing on the network structures that emerge from these interactions (For details on the *Netscan* project see: Smith, 1999; Turner, Smith, Fisher, & Welser, 2005). However, *Netscan* did not include the actual message content. A team of researchers at Carnegie Mellon University was able to capture the original archived message content by retrieving the archived data from *Google Groups* for a random selection of 99 groups from the *Netscan* data from June 2003 to January 2005 (Wang, Kraut, Butler, Burke, & Joyce, under review).

From this original data set (i.e., 2.2 million messages from 99 discussion groups), I randomly sampled a strata of four types of discussion groups: (a) Politics; (b) Health and Support (c) Recreation and Hobby; and (d) Science and Technology, which researchers argue are among the most popular types of online community topics (Horrigan, 2001). From these four categories I sampled sixteen groups using a random integer generator. This results in a sample of messages over a twenty-month period, from June 21, 2003 through January 31, 2005. Table 3-1 lists the names and descriptions of the randomly selected discussion groups.

Table 3-1. *Summary of Google Group Names, Categories and Descriptions*

Name	Category	Description
alt.politics.economics	Politics	War == Poverty, & other discussions

alt.politics.radical-left	Politics	Who remains after the radicals left?
alt.politics.usa.constitution.gun-rights	Politics	Constitutional ramifications of gun rights
talk.politics	Politics	[General political discussions]
alt.support.cancer.breast	Health and Support	Support for those diagnosed with breast cancer and their families
alt.support.depression	Health and Support	Depression and mood disorders
alt.support.diabetes	Health and Support	Support for dealing with diabetes and related topics
alt.support.hepatitis-c	Health and Support	[Support for dealing with hepatitis-c and related topics]
rec.arts.manga	Recreation and Hobbies	All aspects of the Japanese storytelling art form.
rec.crafts.textiles.quilting	Recreation and Hobbies	All about quilts and other quilted items
rec.music.blueNote:blues	Recreation and Hobbies	The Blues in all forms and all aspects
rec.food.veg.cooking	Recreation and Hobbies	Vegetarian recipes, cooking, nutrition
sci.op-research	Science and Technology	Research, teaching & application of operations research.
sci.chemistry	Science and Technology	Chemistry and related sciences.
sci.lang	Science and Technology	Natural languages, communication, etc.
microsoft.public.security	Science and Technology	Deals with security issues for

	Microsoft products

Note: Descriptions and categories were identified from information from GOOGLE GROUPS. In some cases, the description was ascertained after reviewing several discussion group messages. These are marked with brackets ("[ ]").

Procedure

All the interactions and message content that took place over the 20-month period were automatically stored in an offsite database (Smith, 2007; Wang, et al., under review). All analyses involved two additional steps. First, I pulled user log data over a specific time period for 16 randomly selected groups. Second, I analyzed the language content of each message contributed by a user. Both steps are discussed in more detail below.

User Logs

All information regarding the communication behaviors of the authors, such as how often they posted, which messages were replies or thread starters, and when they first appeared and left the topic group, are captured as log files on a *mySQL* database. In addition, I relied on the user logs to create the sociomatrices for the social network analyses based on the reply structure of the group. After completing the text analysis (described below), I aggregated the message-level data with the user logs based on the ID of each author to create a new dataset of 38,483 individuals. Additionally, for use in the hierarchical linear model analyses described later, I created a second data set aggregated by the sixteen topic groups, resulting in the final sample size of 33,540.

While basic posting behavior was captured in the original *Netscan* and *CMU* data set, the actual structure of each message thread in the sample was recreated in order to examine social influence. In other words, the parent and child messages connected to each original message were identified for every author (the calculations used in the social influence measures are described below). This resulted in a sub-sample of 33,644 participants.

Word Frequency Analysis

The messages collected from *Google Groups* content were converted to text files in order to be analyzed by text analysis software. The text files were preprocessed before analysis began, and all headers, subject lines, signatures and quoted text were removed to exclude extraneous noise and ensure authenticity of the main content of each message (Hoffmann, 2007). My original dataset consisted of 653,042 observations at the message level. From this set, 138,919 messages were blank. These blank messages may be a result of any binary post in which users attach an image, document or sound file, or of a message that only includes quoted material or otherwise does not contain any original text. After removing these observations, the sample size was 514,023. After completing the text analyses, I merged its results with the user log data by matching the unique message identifiers.

Each text file was then processed using *Linguistic Inquiry and Word Count (LIWC)* (Pennebaker, et al., 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 output of LIWC is a word frequency ratio of the number of words matched to the internal dictionaries with the total number of words used in the texts.

Social Network Analysis

This research also relies on UCINET (Borgatti, Everett, & Freeman, 1999), a software package which calculates social network statistics and draws network graphs to illustrate the relationships of actors in a network, or in this case, participants in an online community. In order to analyze data, the participants of the community must be identified in terms of a sociomatrix, which identifies any links between the participants. These sociomatrices can be represented in two ways: as a directed graph, in which the values are continuous and asymmetric; or as an undirected, symmetric set of absent (0) or present (1) values (Hanneman & Riddle, 2005). Some social network statistical evaluations require symmetric graphs (i.e., reciprocity centrality), while others allow asymmetric values (i.e., indegree, outdegree and betweenness centrality).

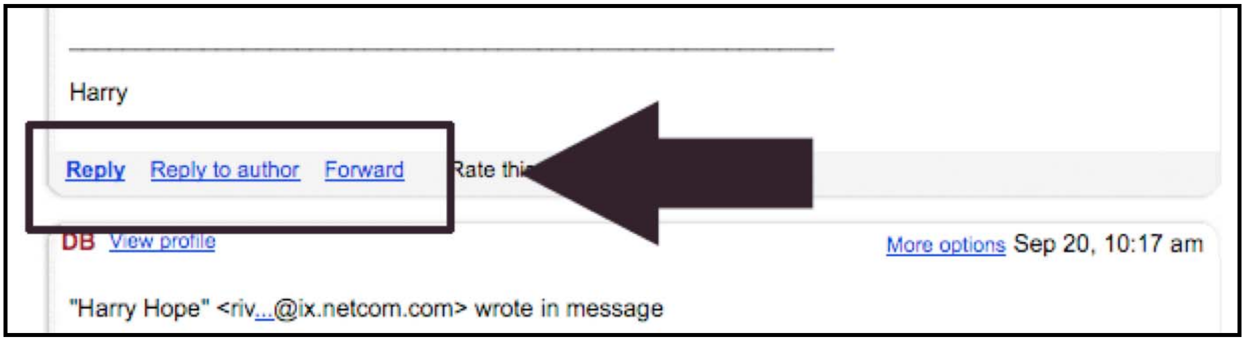
Table 3-2 is provided as an example of a sociomatrix. Participants are labeled *A* through *G* and the rows represent the senders, while the columns represent the targets. The columns represent users who receive a message from a user in each row. This creates a variety of possibilities. For example, while *H* sends messages to everyone, they are not reciprocated by other members of the network. Vice versa, *G* receives messages but does not send messages to anyone. I should also note that many sociomatrices could be sparse and filled with zeros, suggesting that many users do not interact with all other members of the community.

Table 3-2. *Example of Sociomatrix*

	A	B	C	D	E	F	G	H
A	—	4	0	12	15	1	0	0
B	23	—	0	1	0	0	0	0
C	56	0	—	0	0	0	9	0
D	0	0	0	—	0	0	1	0
E	0	0	0	0	—	0	1	0
F	0	0	0	0	0	—	1	0
G	0	0	0	0	0	0	—	0
H	10	12	15	9	10	11	8	—

In order to prepare the data for these analyses, a matrix was created for each discussion group using PHP/mySQL scripts that examine the database and export text files based on the links between each participant. Each message in the database is associated with a unique identification number, as well as an ID associating it as a new message, as a new post and start of a thread, or as a reply to specific author. As shown in *Figure 3-2*, each message offers three response options. A user can reply to a specific author or to the broader discussion group, which starts a new thread. Additionally, a user can create a new post without being part of any other thread.

Figure 3-2. Example of reply options for each message.



Because the *Netscan* database automatically determined if a message was new, or a reply to a specific user, I was able to identify whenever one user is responding to another. Listing every user that has contributed to the group creates the sociomatrices. When a user replies to a message, he becomes the sender. The author of the initial message becomes the target, and receives a positive score for each reply. This allowed me to create a directed graph with continuous values representing the actual number of responses that each participant received from every other participant in the community.

Dependent Measures

Leadership

Again, I define online leadership in terms of the ability to influence other members of the group. This can occur in one of three ways: (a) as a *Reply Trigger*, or the ability to inspire users to respond to posts; (b) through *conversation creation*, or the amount of lengthy conversations that users inspire; and (c) *language diffusion*, or the extent to which the words or topics that users include in messages diffuse along a conversational thread. I describe each of these in more detail below.

All three dependent variables represent count data; it is a non-negative and integer-based variable that captures how often an event occurs (in this case, posting or reply behavior) (Cameron & Trivedi, 1998). Count data is typically a positive skew, representing a non-normal distribution, which is the case with these leadership variables. I rely on a Poisson regression model to examine the influence of the predictors on leadership (this is discussed in more detail in the next section).

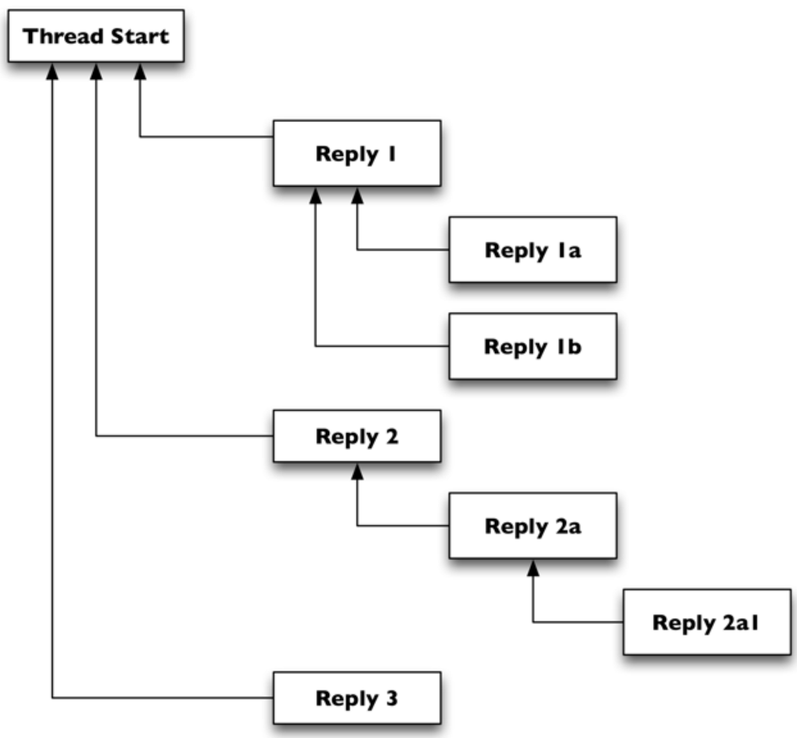
Reply Trigger

The ability to inspire users to reply to message is measured in terms of indegree centrality, which represents the number of incoming links to any participant in the community. In this case, it is measured by the number of responses that a participant receives from other participants in the community. More formally, it is calculated as the sum of the links terminating at actor n_i (Wasserman & Faust, 1994, p. 126).

Conversation Creation

Social influence occurs when users post a message or reply that sparks a long dialogue between other users (Matsumura & Sasaki, 2007). *Figure 3-3* illustrates how an initial post branches out into a series of conversational threads. *Author A* starts a thread about a new health product and receives three direct replies. *Reply 1* and *2* spur additional replies, and *Reply 2a1* is a third-tier reply. Because *Author A* initiated the discussion or question, he directly or indirectly influences each of these tiers. The sum of all scores for all authors who post a message or reply are calculated for each user.

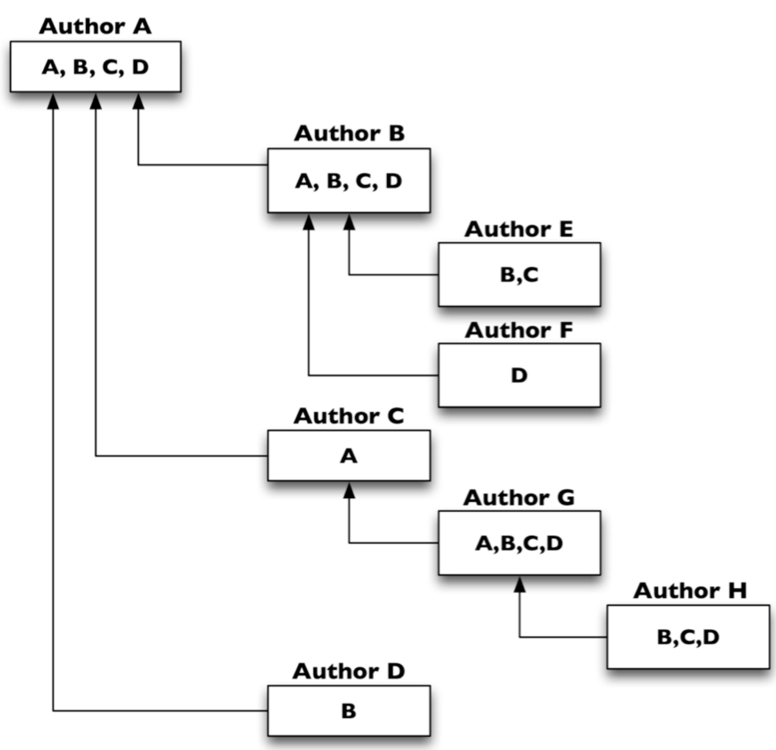
Figure 3-3. Example of reply hierarchy in a threaded discussion.



Language Diffusion

Social influence follows a similar pattern when including language. After identifying which messages are connected in the thread, I examine the number of shared words between them. 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 A* can be considered even more influential since she is farther away from *Author C*. Opposite to the weighting procedure used for conversation creation, I actually double the weight for each word that is shared as it gets further away from the source of the message chain. This is illustrated in *Figure 3-4*; four terms (A, B, C, D) were found in both the starting message (i.e., *Author A*) and the first reply (i.e., *Author B*). Only one term in the starting message was found in the second and third replies (i.e., *Authors E* and *F*). Further in the message chain, two terms from the starting message (B, C) were found in the second tier reply. Language diffusion is calculated as the sum of the shared words between all messages in the chain.

Figure 3-4. Example of language diffusion across a threaded discussion.



In order to capture the diffusion of words, I rely on `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 normalized by the length of the each file. This ensures that longer messages do not have a greater chance of diffusing words since there are more opportunities.

In order to make sure that the words that are measured are meaningful, I pre-processed the messages a second time to remove any stop-words (i.e., a, an, the, etc.). This way simple words such as 'the' are not considered influential and do not muddle the measure. The list of stop words that were removed is available in the Appendix. A second script was created, which counted the total number of words that a participant shared across all interactions with other users, taking into account the social distance metric, which increases the score as a participant's shared words are reflected further down a message chain.

Independent Measures

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 pre-defined dictionaries outlined by Pennebaker, Booth & Francis (2006). For each dimension, the frequency of words captured by the dictionaries is divided by the total words in the message, creating a ratio that controls for the size of the message. In order to aggregate the data from message-level to individual-level, the mean score of each dimension is calculated for each individual. In addition, each linguistic dimension was log-transformed to reduce the positive skew in the data distribution.

Talkativeness

This is measured as the average length of messages contributed by each participant. The sum of the total words found in each message by an author was captured, and then divided by the total number messages, creating an average message length.

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. There are 915 words used in the 'Affective Processes' dictionary. The complete list of words can be found in Appendix B.

Assertiveness

This is measured as the frequency of certainty words such as 'always' or 'never' used by each participant. There are 83 words used in the 'Certainty' dictionary. The complete list of words can be found in Appendix B.

Linguistic diversity

I measure linguistic diversity in terms of lexical diversity and vocabulary richness (Bradac, et al., 1979). This is measured as the number of unique words found in a message, calculated by dividing the number of number of different words (e.g., types) by the total number of words (e.g., tokens). This is featured in (1):

$$\#UniqueWords = {}^{Type}\!/_{Token} \quad (1)$$

Some scholars argue that the type/token ratio (TTR) is influenced by the number of words in the sample, so that larger samples of language can lessen the value of the TTR as the number of unique words introduced in a text steadily diminishes (Malvern, Richards, Chipere, & Durán, 2004; McKee, Malvern, & Richards, 2000). However, given the small size of the average message in this online community, the TTR should not be deflated.

Measurement of Social Networks

In order to examine the popularity or expansiveness of each participant, as well as his value as an intermediary in the network, I rely on four measures of centrality outlined by Freeman (1979) and Wasserman & Faust (1994). Sociomatrices were captured by identifying the reply-structure of each message in the group over the 20-month period. If a message is a reply, then the target message is given an incremental value of "1". For any two messages that are not connected, that cell received a "0". After the connections were identified for all messages, the matrix was then collapsed to represent the actual authors. This results in sixteen matrices (Author × Author) that represent the number of times that an author replied to another author.

Expansiveness

Expansiveness is measured in terms of outdegree centrality, which represents the number of outgoing links from any participant in the community. In this case, it is measured by the number of responses that a participant posts to other participants in the community. It is calculated as the number of links originating from node n_i (Wasserman &

Faust, 1994, p. 126). This measure was log-transformed to reduce the positive skew in the data distribution.

Reciprocity

Reciprocity is measured as the frequency of an individual's participation in a mutual dyad (Wasserman & Faust, 1994). A mutual dyad is one in which both actors reply to one another, regardless of the order of the replies. In order to calculate this, I convert the networks into symmetric relations such that a present tie only appears for each dyad if both replied to one another. Reciprocity is calculated in the same manner as indegree. This measure was log-transformed to reduce the positive skew in the data distribution.

Brokering

Brokering is measured in terms of betweenness centrality, which represents a participant's intermediary value to other members of the network based on their ability to spread and control information and communication flow (Degenne & Forse, 1999). Betweenness is measured by examining the shortest paths, or geodesics, in the network, and determining which participants link these geodesics. 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:

$$C_B(n_i) = \sum_{j<k} g_{jk}(n_i)/g_{jk} \tag{2}$$

In (2), g_{jk} represents the number of geodesics linking actors *j* and *k*, and $g_{jk}(n_i)$ is the number of geodesics linking actors *j* and *k* when containing actor *i* (Freeman, 1979; Wasserman & Faust, 1994). This measure was log-transformed to reduce the positive skew in the data distribution.

Posting Behavior

Number of Posts

This is calculated as the total number of posts that a participant contributes to his respective topic groups.

Number of Replies

This is calculated as the total number of replies that a participant contributes to messages posted by other members of their respective topic groups.

Tenure in Community

This is calculated as the total number of days that a participant actively posts a message or replies to a message in the community.

Measurement of Group-Level Attributes

By combining the user log data, I am able to capture unique characteristics of each topic group, including its membership size and the dynamics of overall participation. This creates a nested structure, providing several group-level predictors to examine leadership.

Except for group stability, all group variables were log-transformed to reduce the positive skew in the data distribution.

Group Size

Group size was calculated as the average number of authors that contribute to the topic group every three months for the twenty-month period. This was chosen rather than the total number of authors over the same period, since the rate of attrition is high and users should only be influenced by the size of the community by which they are surrounded.

Network Density

Network density represents the number of connections between all dyads in a network. In simple form, it is calculated as the proportion of connections between authors against the total potential connections available. In directed graphs, density can be interpreted as the average strength of ties among participants (Hanneman & Riddle, 2005). In this case, density is calculated as:

$$\Delta = \frac{L}{g(g-1)} \tag{3}$$

In (3) L is the number of arcs, or ordered pairs of nodes, and $g(g - 1)$ is the possible number of arcs in the network (Wasserman & Faust, 1994).

Participation Equality

Participation inequality within each group represents the distribution of the proportion of participation, whether a message post or a reply, by all members of a particular topic group. Smith (1999) proposed a poster-to-post ratio to measure interaction quality in each group; however, this measure does not capture reply structures as well. Therefore, I propose a different measure of interaction quality that takes into account all manner of participation.

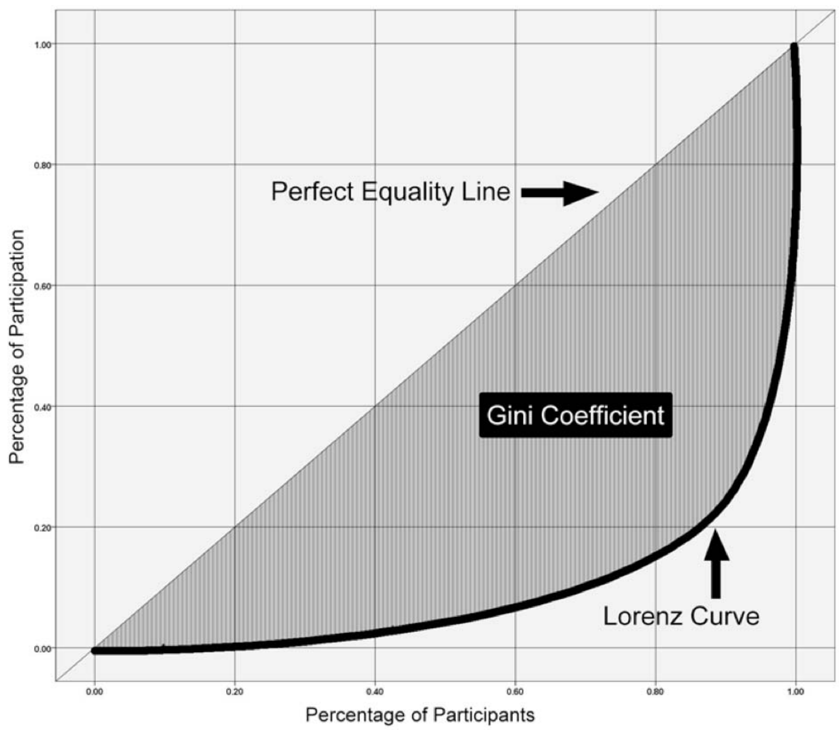
In this study, participation equality is calculated using a Lorenz curve and the inverse of the Gini coefficient. The Lorenz curve, typically used to graphically show the inequality of an income distribution (Kakwani, 1977), demonstrates the proportion of total income given a percentage of a population. An example of this is Pareto's argument that 80% of the wealth resides with 20% of the Italian population (Rosen & Resnick, 1979).

This concept is applied here to represent a cumulative distribution of participation within each topic group. As shown in *Figure 3-5*, the x-axis represents the cumulative proportion of participants ranked by their participation level, and the y-axis represents the cumulative proportion of participation for a given proportion of the user population. For a technical description of the Lorenz curve calculation, see Gastwirth (1972). To measure participation, I calculated the total posts and replies for each participant.

After a Lorenz curve is calculated, the degree of inequality can be measured using the Gini coefficient (Gastwirth, 1972). First, a line of perfect equality is created where y = x (i.e., a 45 degree line). Second, the Gini coefficient is calculated as a ratio of: (a) the area between the line of perfect equality and the Lorenz curve, and (b) the area beneath the Lorenz curve and the axes (i.e., a line of perfect inequality). The higher this coefficient is, the more *unequal* the participation distribution remains. For a technical description of the Gini coefficient calculation, see Atkinson (1970).

Because the Gini coefficient represents inequality, the inverse of the Gini coefficient was calculated in order to represent participation equality.

Figure 3-5. Example of a Lorenz curve, perfect equality line and Gini coefficient for a random topic group.



Group Stability

Group stability was measured as the absolute value of the average percentage change of contributing authors every three months over the 20-month period.

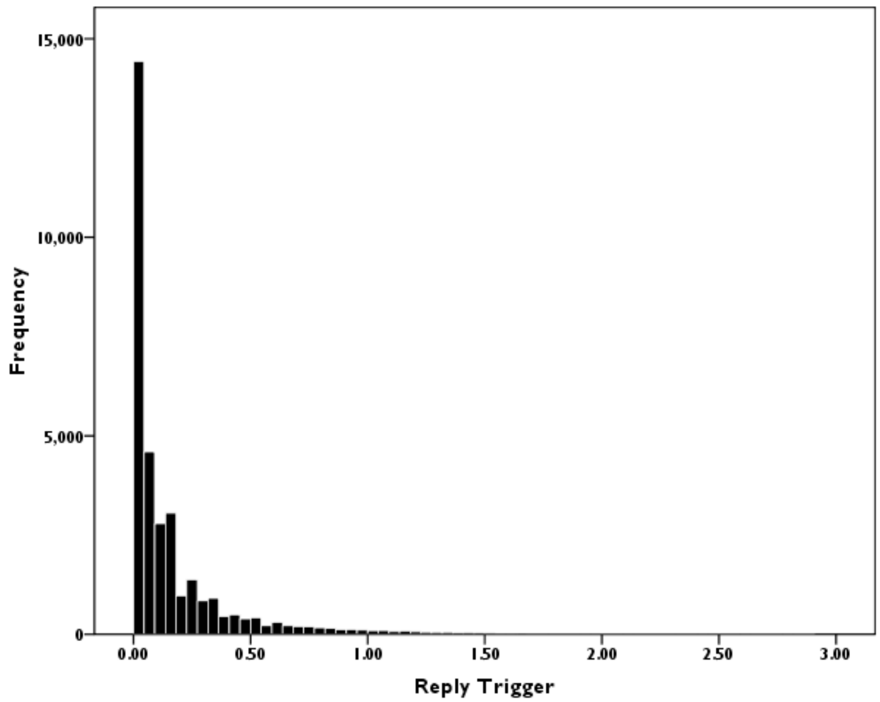
$$\sum \frac{\left|t_{n+1} - t_n\right|}{t_n} \tag{4}$$

In (4), *t* is the average number of authors, and *n* is a particular time period. Then we take an average across all time periods to create a final percent change value, which can be positive or negative.

Chapter 4. Results

In order to examine the relationship between emergent leadership, social networks, language use, diffusion and group influence, I rely primarily on hierarchical linear modeling. As previously discussed, the dependent variable represents event count data, which requires special generalized modeling approaches (see Figure 4-1). Traditional linear regression models require assumptions that are not evident in count variables, and often result in erroneous predictions (Gardner, Mulvey, & Shaw, 1995). The benchmark of the nonlinear alternative is the Poisson, or log-linear regression model, which seeks to take advantage of the non-negative and integer-based aspects of the outcome variable (Cameron & Trivedi, 1998). Note that Poisson regression relies on a log-transformation of the dependent variable, and requires an antilog transformation of the coefficients of each predictor in the regression model (Gardner, et al., 1995; Gelman & Hill, 2007). Throughout these analyses, I will also include the *exp(B)* — the antilog — when interpreting each variable.

Figure 4-1. Histogram of online leadership



Note: This dependent variable was log-transformed to show the distribution more clearly.

Descriptive Statistics

In this section, I present a general description of the data and variables of interest. I also examine extreme values and cross-posters to distinguish spammers or other idiosyncratic behaviors that are not representative of the sample under study.

Means and Correlations for Individual-level Variables

The mean, standard deviation and range for each individual-level variable of interest are presented in Table 4-1. The majority of the variables fall on a scale within the hundreds, but Talkativeness and the social network measures all have a higher range. Betweenness, which is defined as the number of times a participants connects to different nodes that do no connect with each other, is on a much larger scale with a maximum of over five million.

Table 4-1. *Summary of Means, Standard Deviations and Range for all Variables of Interest*

Variables	*M*	*SD*	*Min*	*Max*
Dependent Variables				
Reply Trigger	14.05	93.24	0	7,369
Conversation Creation	12.08	86.04	0	5,945.36
Language Diffusion	1.05	10.14	0	731.21
Independent Variables				
Number of Posts	20.95	147.00	1	11,174
Number of Replies	18.36	126.95	0	6,663
Tenure	7.31	26.88	1	578
Expansiveness	14.06	96.24	0	5,945
Reciprocity	8.04	65.46	0	4,545
Brokering	9,315.98	94,329.70	0	5,345,361
Talkativeness	181.67	752.18	1	26,230
Affect	5.27	4.86	0	100
Assertiveness	1.39	2.19	0	100
Linguistic Diversity	79.88	15.09	.32	150

Note: N = 33,540.

A correlation analysis of all variables of interest is presented in Table 4-2. 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).

There are small, positive correlations between the number of posts and the number of replies (r = .11, p <.001), and the posts and tenure in the community (r = .23, p <.001). There is a small correlation between expansiveness (outdegree) and brokering (betweenness) (r = .38, p <.001) and a strong correlation between expansiveness and reciprocity (r = .97, p <.001). Collinearity tests suggest that both outdegree (Tolerance = .05, Variance Inflation Factor (VIF) = 21.71) and reciprocity (Tolerance = .04, VIF = 27.12) are possibly redundant variables (Hair, Anderson, Tatham, & Black, 1998). Reciprocity, which is calculated as the number of reciprocal links, is comprised of both the indegree and outdegree variables, so a collinearity issue makes sense.

Talkativeness and linguistic diversity are negatively correlated (r = –.37, p <.001), which resonates with previous arguments that the type-token ratio decreases as the number of words increase (Malvern, et al., 2004; McKee, et al., 2000). However, a collinearity test does not reveal a problem with these variables since Tolerance is much greater than 0 and

VIF is less than 20[7] (Talkativeness Tolerance = .83, VIF = 1.2; Linguistic Diversity Tolerance = .83, VIF = 1.21).

There are small, but significant correlations between affect and assertiveness (r = .07, p <.001), and linguistic diversity (r = .07, p <.001). This is likely because some emotional words can appear in the same dictionaries for assertive language and all words are calculated in the type/token ratio that represents linguistic diversity.

Means and Correlations for Group-level Variables

The means, standard deviations and range of each group variable are listed in Table 4-3. Size represents the average number of authors participating in the group during each quarter of the 20-month time period (M = 362.58, SD = 353.49). The other group variables, participation equality and network density, suggest that these groups are pretty sparse and unequal in the amount of participation among all authors. Turnover, which represents the percent change of authors during all eight quarters of the 20-month period, suggests a general increase in authors over time (M = .20, SD = .93). Again, many of these variables utilize a different scale of measurement, which is addressed in the regression analyses by transforming some variables to the same scale.

---

[7] "0" and "20" are widely accepted tolerance points for collinearity tests. See Menard (1995) for more on information.

Table 4-2. *Correlation Matrix for all Variables of Interest*

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Reply Trigger	—	.965**	.841**	.514**	.811**	.749**	.863**	.388**	.891**	–.007	.016**	.002	.015**
2. Conversation Creation		—	.857**	.487**	.795**	.708**	.856**	.387**	.888**	–.007	.012*	.003	.018**
3. Language Diffusion			—	.304**	.744**	.696**	.810**	.328**	.835**	–.007	.007	.000	.009
4. Number of Posts				—	.108**	.232**	.104**	.269**	.108**	.006	.001	-.004	–.014**
5. Number of Replies					—	.811**	.940**	.337**	.908**	–.011*	.021**	.003	.029**
6. Tenure						—	.780**	.412**	.730**	-.009	.018**	-.001	.016**
7. Expansiveness							—	.376**	.972**	–.011*	.020**	.004	.031**
8. Brokering								—	.295**	–.001	.003	.003	–.001
9. Reciprocity									—	–.010	.018**	.004	.026**
10. Talkativeness										—	.003	.005	–.369**
11. Affect											—	.074**	.071**
12. Assertiveness												—	-.005
13. Linguistic Diversity													—

Note: Bivariate correlations, two-tailed tests. N = 33,540. *p <.05. **p <

As shown in Table 4-4, there were no significant correlations between the group-level predictors. There are also no significant differences between group type (i.e., politics vs. health vs. hobbies vs. science) and the group-level variables. A chi-square shows no significant differences for size (p = .35), participation equality (p = .52), density (p = .43) or turnover rate (p = .35). So while there is variance across all sixteen groups, no topic area shows a significantly different size, complexity or stability.

Table 4-3. *Summary of Means, Standard Deviations and Range for Group-Level Variables of Interest*

Variables	*M*	*SD*	*Min*	*Max*
Size	362.58	353.49	15.13	1065.88
Participation Equality	.21	.10	.07	.42
Network Density	.008	.0009	.0002	.034
Turnover	.55	.99	.12	4.23

Note: N = 16.

Table 4-4. *Correlation Matrix for Group Size, Complexity and Turnover*

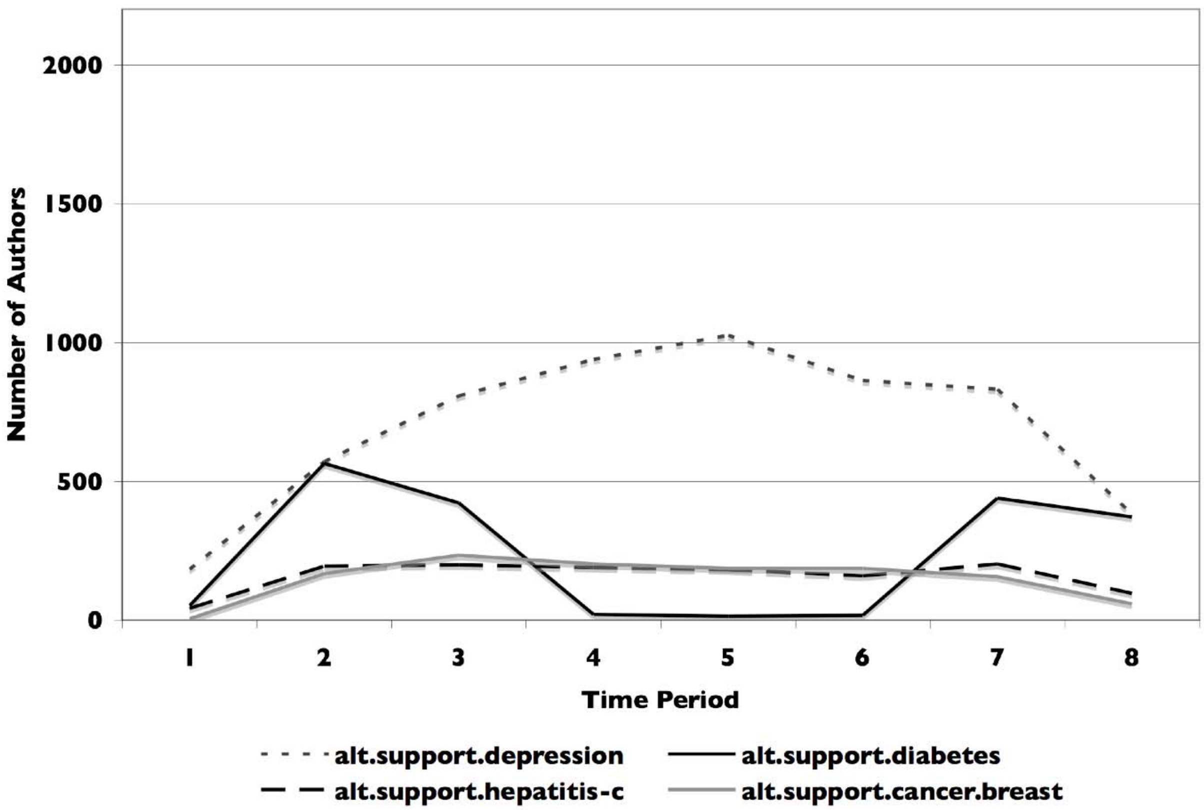
Variable	1	2	3
1. Size	—		
2. Participation Equality	–.493	—	
3. Network Density	–.331	–.307	—
4. Turnover	–.060	–.174	.050

Note: Bivariate correlations, two-tailed tests. N = 16.

A Note on the Dynamics (and Attrition) of Online Groups

There is a high attrition rate associated with many online communities (Andrews, Nonnecke, & Preece, 2003), of which these online discussion groups are no exception. In fact, in this sample, only 56 individuals (less than 1%) contribute over the entire 20 months. Of these, 21 individuals are from *alt.support.depression*, while 16 reside in *talk.politics*. Six of the sixteen groups have no authors that remain through the entirety). In the groups focused on hobbies and recreation, no one contributed over the entire 20-months, while the majority of long-term contributors came from groups focused on health and support (i.e., 25 participants) and politics (i.e., 22 participants), with science and technology a distance third (i.e., 9 participants). *Figure 4-2* illustrates membership size and change over the 20-month period, organized by genre. It shows that some groups have a relatively stable group of members while others show massive peaks and valleys.

Figure 4-2. An example of membership size and turnover in four health and support discussion groups.



In fact, the size and contribution rates of this random sample of discussion groups vary. At least four groups show higher frequencies of posting new messages and replying to other messages. It is also important to note that these prolific groups are not always the ones with the most authors. For example, *alt.support.diabetes* and *rec.crafts.textiles.quilting* have fewer authors and still contribute at the highest levels. By contrast, *microsoft.public.security* has many authors, but not many contributions. See *Figure 4-3* and *Figure 4-4*.

A Note on Extreme Values and Cross Posters

Given previous work that suggests that spammers are prevalent in these communities (Smith, 1999; Turner, et al., 2005), I examined the extreme cases in my sample to look for instances of spam behavior. I took the top 30 contributors and reviewed their incoming and outgoing links, as well as the actual message content of a random sample of their messages. Turner et al. (2005) argue that spammers can be identified as having large outgoing links with no incoming ones. However, among the top 30 contributors, there is symmetry between these two centrality values. This also provides some insight into why the correlation between indegree and outdegree is so high.

As I began to randomly sample the message content for each leader, I noticed that some messages were blank, indicating either a binary post, or a post that contained no original text (such as quoted text only) or an error in the data scrubbing. While these observations were removed for the final analysis (see Procedure sub-section for more information), they are also good indicators for spamming or other nonhuman behaviors (Smith, 1999). Therefore, I used this as an additional indicator. By combining these methods, I only note one author (coincidently, the highest poster) that was out of the ordinary. First, 13,741 of the collection of messages were blank. A random sample of content revealed incomprehensible text, messages in foreign languages, and advertisements. This author was removed from the final analysis. The other top leaders did not show abnormalities in their network links, blank messages or in the content of each message.

Figure 4-3. Post and reply behavior of 16 randomly selected discussion groups.

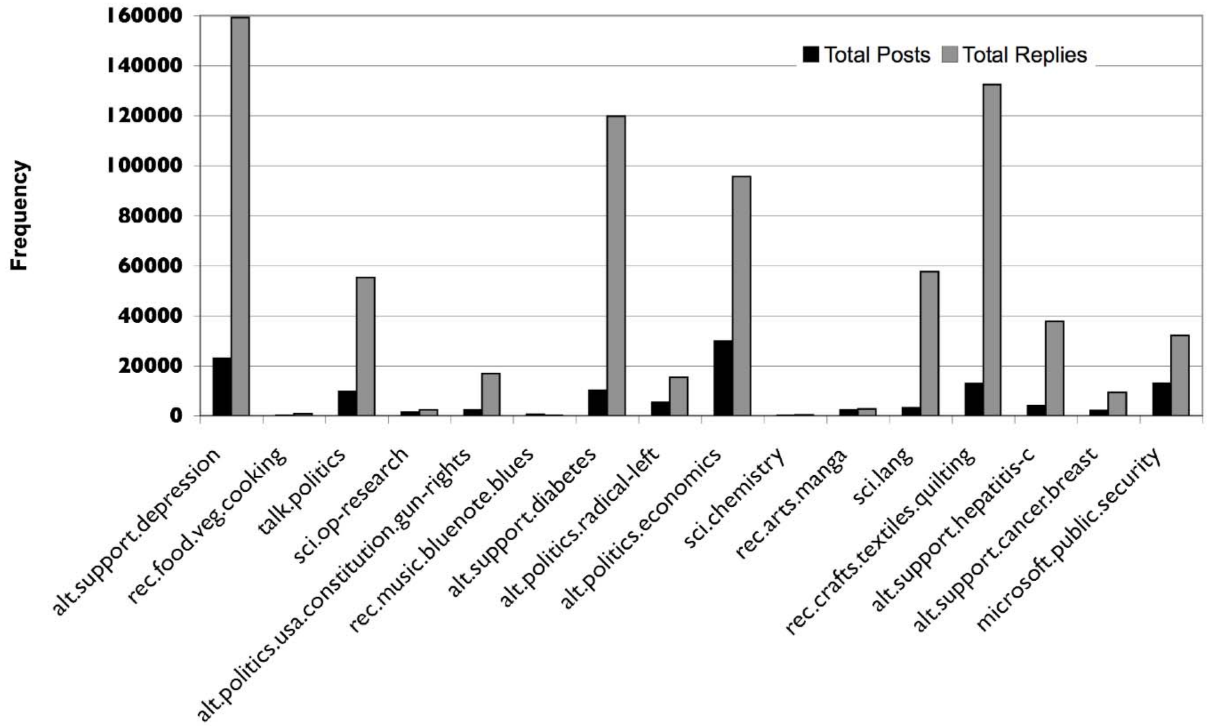
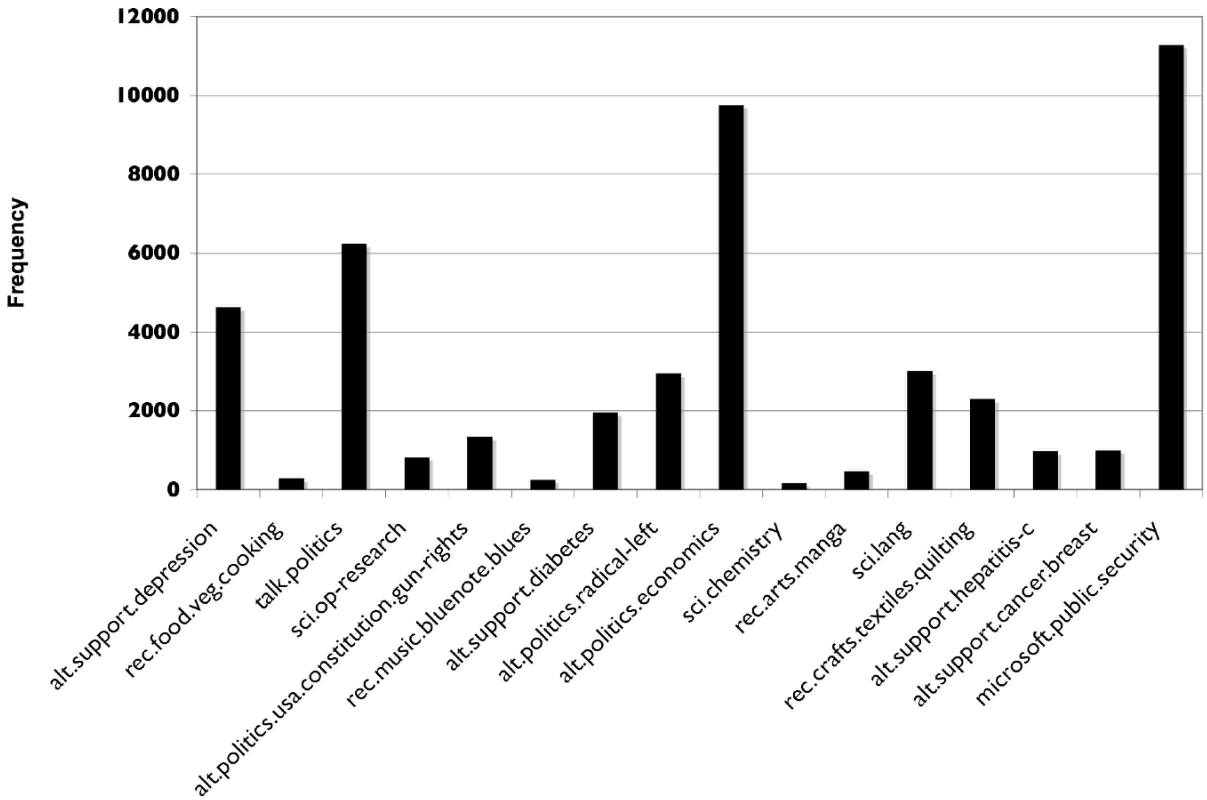


Figure 4-4. Total number of authors that contributed to 16 randomly selected discussion groups.



A Note on Extreme Values and Cross Posters

Given previous work that suggests that spammers are prevalent in these communities (Smith, 1999; Turner, et al., 2005), I examined the extreme cases in my sample to look for instances of spam behavior. I took the top 30 contributors and reviewed their incoming and outgoing links, as well as the actual message content of a random sample of their messages. Turner et al. (2005) argue that spammers can be identified as having large outgoing links with no incoming ones. However, among the top 30 contributors, there is symmetry between these two centrality values. This also provides some insight into why the correlation between indegree and outdegree is so high.

As I began to randomly sample the message content for each leader, I noticed that some messages were blank, indicating either a binary post, or a post that contained no original text (such as quoted text only) or an error in the data scrubbing. While these observations were removed for the final analysis (see *Procedure* sub-section for more information), they are also good indicators for spamming or other nonhuman behaviors (Smith, 1999). Therefore, I used this is an additional indicator. By combining these methods, I only note one author (coincidently, the highest poster) that was out of the ordinary. First, 13,741 of the author's messages were blank. A random sample of content revealed incomprehensible text, messages in foreign languages, and advertisements. This author was removed from the final analysis. The other top leaders did not show abnormalities in their network links, blank messages or in the content of each message.

There also appears to be a pattern of many of the top leaders coming from the same groups. For example, eleven of the top leaders participate in alt.support.depression (i.e., *Group 816*). Another health-related group, alt.support.diabetes (i.e., *Group 9386*) contains seven of the top leaders, and rec.crafts.textiles.quilting (i.e, *Group 26772*) contains six of the top leaders. The rest of the leaders were distributed evenly among the rest of the groups, although I should note that all the top leaders participate in only nine of the sixteen groups used in this study.

Several of the top leaders are also cross-posters, although in most cases they cross-post to only one other group and at a much smaller level (ranging from four to 60 cross-posts). However, subsequent analysis did not reveal a significant relationship between cross-posting and leadership ($p > .50$). However, it is important to note the level of cross-posting that occurs in these discussion groups. Overall, there were 4,234 cross-posters in the sample, representing 7,354 observations in the sample. This was considerably smaller than the cross-posting behavior that Smith (1999) observed in his initial study. In order to account for the group-level variance, each participant that cross-posted received a unique observation point. This allows me to determine if leaders utilize different traits or interactions when cross-posting to different groups.

A Note on Question-based Posts

Online communities are often conceived as a place to ask questions. Some communities such as YAHOO! ANSWERS are dedicated to this, but scholars have studied this phenomenon in online discussion groups and determined that it is a primary activity (Turner, et al., 2005). Here it might seem intuitive that those who trigger a reply or spark a conversation may do so by starting with a question. If this were the case, they aren't necessarily influencing others.

However, regression analysis did not reveal any correlation between asking questions (or at least, ending statements with a "?") and leadership as a reply trigger (p = .78), as conversation creation (p = .99) or as language diffusion (p = .70). Therefore, while asking questions may be a common activity, it is likely that it comes from users who ask a question and leave, not from an online leader.

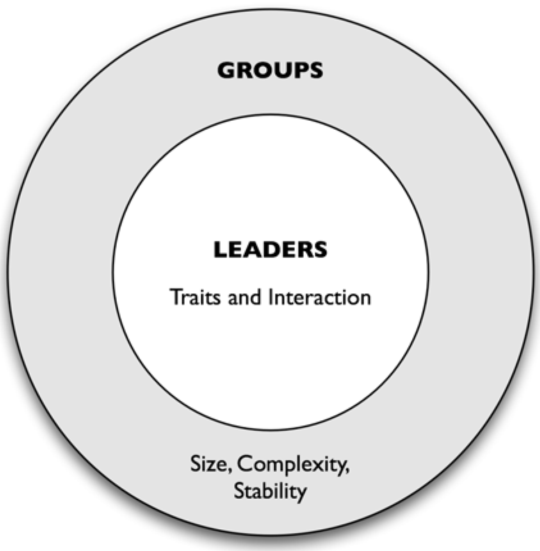
Summary

The descriptive statistics show that many of the variables of interest — including leadership —represent positive skews. This is not surprising since research has shown that the majority of contribution is done by a handful of members. What is surprising is that highest levels of leadership tend to come from a handful of groups (such as health-related groups). Second, the data reflects the dynamic nature of online communities. There is a high rate of attrition and most members do not stay long. Similarly, the size of groups and amount of participation varies dramatically, reinforcing the need for multi-level modeling approaches.

Hierarchical Linear Modeling and the Poisson Distribution

In order to test my hypotheses regarding the traits and interaction styles associated with emergent leaders in online discussion groups, I utilize hierarchical linear modeling (HLM). Again, the data set represents a nested structure; that is, while message content is attributed to individual participants, they are embedded in the larger organizational culture of the topic group, which has its own set of unique characteristics. This relationship is depicted in *Figure 4-5*, where the individual variables related to leadership are embedded in topic groups that have their own size, complexity and stability variables. HLM also allows me to measure the level of interdependence between the individual-level and group-level variables, as well as their independent effects on leadership. For all the models presented in this section, I rely on HLM 6.0, a statistics program devoted exclusively to multi-level models (Raudenbush, 2004).

Figure 4-5. Example of contextual factors in which leaders are embedded in groups.



Because the dependent variable represents count data, I utilize Poisson regression, which is the benchmark for modeling count data (Cameron & Trivedi, 1998). In a Poisson distribution, it is expected that the variance equals the mean; when the variance is larger than the mean, it is referred to as overdispersion and this occurs quite often (Gelman & Hill, 2007). Such is the case in my data set, in which chi-square results far exceed the degree of freedom (a simple measure of overdispersion), so I adjust all models for overdispersion. The parameter estimates do not change in magnitude or direction when adjusting for overdispersion.

In addition, I rescaled the social network variables and talkativeness to keep all variables on the same hundredths scale. This type of standardization does not affect the positions of the data points, but does allow the coefficients to be interpreted on the same scale (Gelman & Hill, 2007). This was necessary because HLM 6.0 does not report standardized betas, requiring standardization and normalization before fitting the model (Raudenbush, 2004). I also log-transformed all predictor variables to reduce the positive skew. This creates a log-log model, which can be interpreted in terms of elasticity: if a coefficient is .6, then a 1% change in X creates a .6% change in Y (Gelman & Hill, 2007). Finally, all predictors were grand-mean centered, which is a common and recommended practice in HLM models (Gelman & Hill, 2007), including those dealing with organizations or communities (Hofmann & Gavin, 1998).

Leadership as a Reply Trigger

Baseline Model and Intraclass Correlation

It is typical to begin with a baseline or null model in which none of the predictors are included, allowing an investigator to examine the independence of both the individual- and group-level units on the outcome variable (Hoxx, 1995). For the former, the intraclass correlation (ICC) estimate is used (Hayes, 2006), and calculated as the variance divided by the sum of the variance (τ^2) and the estimated residual variance (σ^2) of the random effects. For the latter, the chi-square difference test is used.

In this case, the ICC is .002, which suggests that less than 1% of the variance in leadership is between-groups and 99% of the variance is at the individual level. For the group-level predictors, the chi-square difference test is used, and the baseline model suggests that differences in discussion group size, equality of participation, network density and turnover rate are different than zero across groups (χ^2 (15) = 1147.85, p <.001).

The Sociability and Trustworthiness of Reply Triggers

After assessing the baseline model, I include individual-level predictors to develop random coefficients models, which allows me to test my hypotheses. The model is a good fit, χ^2 (15) = 1318.54, p <.001, with a Deviance of 154,981.92. While HLM computes a log-likelihood function, which can be converted into the Deviance score, it is possible to determine a proportion of variance explained similar to R^2 found in OLS regression[8]. In this model, the R^2 is calculated by subtracting the σ^2 of the random coefficients model from the σ^2 of the intercept-only model and divide by the σ^2 of the intercept-only model. This results in an R^2 of 0.99, suggesting that sociability and trustworthiness account for 99% of the within-subjects variance.

I predicted that leaders are more likely to post messages (H1a) and reply to the messages (H1b) of other members of the group, and that leaders are more likely to have a longer tenure (H1c) in the group. All three hypotheses are supported. As shown in Table 4-5, the number of posts is positively associated with the number of replies triggered by a leader (i.e, 'reply triggers') (B = .25, SE = .03, $exp(B)$ = 1.28, t(15) = 8.06, p <.001). The number of replies is positively associated with reply triggers (B = 1.24, SE = .20, $exp(B)$ = 3.45, t(15) =

[8] Special thanks to Professor Scott C. Roesch for providing tutorials for calculating R^2 in HLM 6.0. Retrieved from http://www.psychology.sdsu.edu/new-web/FacultyLabs/Roesch/HLM_lab.doc.

6.34, *p* <.001). The tenure, or number of active days in the group is also positively associated with reply triggers (*B* = .73, *SE* = .19, *exp(B)* = 2.08, *t*(15) = 3.81, *p* =.002).

Table 4-5. *Summary of Random Coefficient Regression Analysis of Sociability and Trustworthiness Variables Predicting Leadership as a Reply Trigger*

Variable	*B (SE)*	*SE*	*Exp(B)*
Intercept	1.18***	.08	3.27
Number of Posts	1.24***	.20	3.45
Number of Replies	.25***	.03	1.29
Tenure	.73**	.19	2.08
σ^2	5.90		
τ_{00}	.11		
Deviance	154,981.92		
Chi-Square (df)	1318.54(15)***		

Note: The restricted maximum likelihood method is used for estimation. Predictor variables are estimates of the fixed effects, γs, with robust standard errors, and adjusted for overdispersion. N = 33,540. **p* <.05. ***p* <.01. ****p* <.001.

The Network Centrality of Reply Triggers

I also hypothesized that leaders are more likely to be expansive (H2a), reciprocal (H2b) in their connections with others, and more likely to serve as brokers between otherwise disconnected participants (H2c). The model is a good fit, χ^2 (15) = 1884.34, *p*

<.001, with a Deviance of 188,722.62. This model also shows an R^2 of .96, meaning that the measures of centrality account for 96% of the variance in the number of triggered replies.

As shown in Table 4-6, the first and second hypotheses are supported. Expansiveness, or outdegree centrality, is positively associated with reply triggers (B = 1.23, SE = .37, $exp(B)$ = 3.43, $t(15)$ = 3.31, p =.005). Reciprocity is positively associated (B = 1.84, SE = .13, $exp(B)$ = 6.33, $t(15)$ = 13.86, p <.001). Contrary to prediction, brokering (i.e., betweenness centrality) is not related to reply trigger (p = .64).

Table 4-6. *Summary of Random Coefficient Regression Analysis of Social Network Variables Predicting Leadership as a Reply Trigger*

Variable	*B (SE)*	*SE*	*Exp(B)*
Intercept	1.51***	.12	4.53
Expansiveness	1.23***	.97	3.43
Reciprocity	1.84***	.13	6.33
Brokering	−.17	.35	.84
σ^2	16.15		
τ_{00}	.23		
Deviance	188,722.62		
Chi-Square (df)	1884.34(15)***		

Note: The restricted maximum likelihood method is used for estimation. Predictor variables are estimates of the fixed effects, γs, with robust standard errors, and adjusted for overdispersion. N = 33,540. *p <.05. **p <.01. ***p <.001.

The Language of Reply Triggers

Next, I predicted that leaders would be more likely to demonstrate higher frequencies of talkativeness (H3a), affect (H3b), assertiveness (H3c) and linguistic diversity (H3d). The model is a good fit, χ^2 (15) = 727.52, p <.001, with a Deviance of 293,383. This model also shows an R^2 of .18, meaning that these language variables account for 18% of the variance in the number of triggered replies. All four hypotheses are supported.

As shown in Table 4-7, talkativeness, measured in terms of the average message length, is positively associated with reply triggers (B = 2.42, SE = .14, $exp(B)$ = 11.19, $t(15)$ = 16.85, p <.001). Affect, measured in terms of the relative frequency of emotional words in a message, is positively related to reply triggers (B = .69, SE =.11, $exp(B)$ = 1.99, $t(15)$ = 6.29, p <.001). Assertiveness, measured in terms of the relative frequency of certainty words in a message, is positively related to reply triggers (B = .87, SE =.06, $exp(B)$ = 2.38, $t(15)$ = 15.14, p <.001). Finally, linguistic diversity, measured in terms of the type/token ratio, is positively related to reply triggers (B = 7.05, SE =.83, $exp(B)$ = 1154.07, $t(15)$ = 16.85, p <.001).

Table 4-7. *Summary Of Random Coefficient Regression Analysis For Language Variables Predicting Leadership as a Reply Trigger*

Variable	*B*	*SE*	*Exp(B)*
Intercept	2.28***	.21	9.74
Talkativeness	2.41***	.14	11.19
Affect	.69***	.11	1.99
Assertiveness	.87***	.06	2.38
Linguistic diversity	7.05***	.83	1154.07
σ^2	367.76		
τ_{00}	.65		
Deviance	293,383		
Chi-square (df)	727.52(15)***		

Note: The restricted maximum likelihood method is used for estimation. Predictor variables are estimates of the fixed effects, γs, with robust standard errors, and adjusted for overdispersion. N = 33,540. *p <.05. **p <.01. ***p <.001.

The Effects of Group Attributes on Reply Triggers

Finally, I hypothesized that several group attributes such as the size of the group (H4a), the degree of participation equality (H4b) and the group density or number of connections between members (H4c) would be positively associated with leadership, while the amount of turnover would be negatively associated with leadership (H4d). The means-as-outcomes model, which measures the group-level effect on the individual outcome variable, is a good fit, χ^2 (11) = 39.62, p =.001, with a Deviance of 300,272.20. The group-level R^2 is .94, meaning that group attributes account for 94% of the variance in reply triggers. These hypotheses received partial support.

As shown in Table 4-8, larger discussion groups, measured in terms of the average number of participants every three months for a 20-month period, are positively related to reply triggers (B = 1.79, SE = .30, *exp(B)* = 5.99, t(11) = 6.02, p <.001). Discussion groups with more network density, or more connections between members in terms of message replies, are positively related to reply triggers (B = 1.70, SE = .11, *exp(B)* = 5.49, t(11) = 15.20, p <.001).

Contrary to prediction, discussion groups with more participation equality are not significantly related to reply triggers (p = .91). Nor are groups with a high member turnover related to reply triggers (p = .30).

Table 4-8. *Summary of Hierarchical Regression Analysis for Group Variables Predicting Leadership as a Reply Trigger*

Variable	*B*	*SE*	*Exp(B)*
Intercept	2.23***	.07	12.06
Size	1.78***	.30	5.99
Participation Equality	.06	.53	1.06
Network Density	1.70***	.11	5.49
Turnover	.03	.03	1.03

Variable	B	SE	Exp(B)
σ^2	452.24		
τ_{00}	.04		
Deviance	300,272.20		
Chi-square (df)	39.62(11)**		

Note: The restricted maximum likelihood method is used for estimation. Predictor variables are estimates of the fixed effects, γs, with robust standard errors, and adjusted for overdispersion. N = 16. **p* <.05. ***p* <.01. ****p* <.001.

Summary

The first set of analyses examines the posting behavior, network interactions and language used by online leaders to trigger a reply to the online messages they post. First the findings show that the number of posts, the number of replies and the length of time (i.e., tenure) in the community are all positively related to triggering a reply. Of these, the number of posts appears to have the strongest effect. Second, outgoing links (i.e., expansiveness) and reciprocity are positively related to triggering replies, while betweenness (i.e., brokering) is not related. Of these, reciprocity appears to have the largest effect.

Third, talkativeness, affect, assertiveness and linguistic diversity are all positively related to triggering a reply. Of these, linguistic diversity, which is represented by the number of unique words in a text, shows a considerably higher odds ratio, or effect size. Taken together, the findings support previous research that identifies common factors that increase social influence. That is, communication activity, social networking behaviors, and messages that reflect intensity or complexity increase the chance that an online message will receive a reply.

Fourth, I examined how group attributes influence reply triggers. The results show that group size (i.e., the average number of authors in a three-month period) and group density needs a verb are positively related to reply triggers, while participation equality and membership turnover are not significantly related. This shows that leadership is further influenced by how large and well-connected the group is, suggesting that social influence improves when there are more opportunities to interact.

Leadership as a Conversation Creation

Baseline Model and Intraclass Correlation

In my next analysis, the ICC is .001, which suggests that that less than 1% of the variance in leadership is between-groups and 99% of the variance is at the individual level. For the group-level predictors, the chi-square difference test is used, and the baseline model suggests that differences in discussion group size, equality of participation, network density and turnover rate are different than zero across groups (χ^2 (15) = 768.85, p <.001).

The Sociability and Trustworthiness of Conversation Creation

Again, the model is a good fit, χ^2 (15) = 598.45, p <.001, with a Deviance of 169,455.68. The model also shows an R^2 of .98, suggesting that sociability and trustworthiness account for 98% of the within-subjects variance. Again, all three hypotheses are supported.

As shown in Table 4-9, the number of posts (H1a) is positively associated with conversation creation (*B* = .16, *SE* = .02, *exp(B)* = 1.17, *t(*15) = 7.59, *p* <.001). The number of replies (H1b) is also positively associated to conversation creation (*B* = 1.53, *SE* = .20, *exp(B)* = 4.63, *t(*15) = 7.48, *p* <.001). The tenure (H1c), or number of active days in the group is also positively associated to conversation creation (*B* = .64, *SE* = .21, *exp(B)* = 1.89, *t(*15) = 2.97, *p* =.01).

Table 4-9. *Summary of Random Coefficient Regression Analysis of Sociability and Trustworthiness Variables Predicting Leadership as a Conversation Creation*

Variable	*B (SE)*	*SE*	*Exp(B)*
Intercept	1.09***	.08	3.27
Number of Posts	.16***	.02	1.17
Number of Replies	1.53***	.20	4.63
Tenure	.64**	.21	1.89
σ^2	9.09		
τ_{00}	.11		
Deviance	169,455.68		
Chi-Square (df)	598.46(15)***		

Note: The restricted maximum likelihood method is used for estimation. Predictor variables are estimates of the fixed effects, γs, with robust standard errors, and adjusted for overdispersion. N = 33,540. *p <.05. **p <.01. ***p <.001.

The Network Centrality of Conversation Creation

Again, I hypothesized that leaders are more likely to be expansive (H2a), reciprocal (H2b) in their connections with others, and to serve as brokers between otherwise disconnected participants (H2c). The model is a good fit, χ^2 (15) = 980.66, *p* <.001, with a Deviance of 194,132.30. This model also shows an R^2 of .97, meaning that the measures of centrality account for 97% of the variance in the amount of conversation creation.

As shown in Table 4-10, the first and second hypotheses are supported. Expansiveness, or outdegree centrality, is positively associated with conversation creation (*B* = 2.19, *SE* = .47, *exp(B)* = 8.94, *t*(15) = 4.62, *p* =.005). Reciprocity is positively associated with conversation creation (*B* = 1.33, *SE* = .16, *exp(B)* = 3.81, *t*(15) = 8.17, *p* <.001). Contrary to prediction, brokering (i.e., betweenness centrality) is not related to conversation creation (*p* = .08), although we see a possible trend.

Table 4-10. *Summary of Random Coefficient Regression Analysis of Social Network Variables Predicting Leadership as Conversation Creation*

Variable	*B (SE)*	*SE*	*Exp(B)*
Intercept	1.42***	.11	4.14
Expansiveness	2.19***	.47	8.94
Reciprocity	1.34***	.16	3.81
Brokering	–.67	.36	.51
σ^2	18.99		
τ_{00}	.18		

Variable	B (SE)	SE	Exp(B)
Deviance	194,132.30		
Chi-Square (df)	980.66(15)***		

Note: The restricted maximum likelihood method is used for estimation. Predictor variables are estimates of the fixed effects, γs, with robust standard errors, adjusted for overdispersion. N = 33,540. **p* <.05. ***p* <.01. ****p* <.001.

The Language of Conversation Creation

Next, I predicted that leaders would be more likely to demonstrate higher frequencies of talkativeness (H3a), affect (H3b), assertiveness (H3c) and linguistic diversity (H3d). The model is a good fit, χ^2 (15) = 448.31, p <.001, with a Deviance of 300,072.60. This model also shows an R^2 of .18, meaning that these language variables account for 18% of the variance in the amount of conversation creation. All four hypotheses are supported.

As shown in Table 4-11, talkativeness, measured in terms of the average message length, is positively associated with conversation creation (*B* = 2.44, *SE* = .12, *exp(B)* = 11.51, *t*(15) = 20.37, *p* <.001). Affect, measured in terms of the relative frequency of emotional words in a message, is positively related to conversation creation (*B* = .68, *SE* =.11, *exp(B)* = 1.98, *t*(15) = 6.43, *p* <.001). Assertiveness, measured in terms of the relative frequency of certainty words in a message, is positively related to conversation creation (*B* = .93, *SE* =.06, *exp(B)* = 2.54, *t*(15) = 16.65, *p* <.001). Finally, linguistic diversity, measured in terms of the type/token ratio, is positively related to conversation creation (*B* = 7.62, *SE* =.99, *exp(B)* = 2054.02, *t*(15) = 7.72, *p* <.001).

Table 4-11. *Summary Of Random Coefficient Regression Analysis For Language Variables Predicting Leadership as Conversation Creation*

Variable	*B*	*SE*	*Exp(B)*
Intercept	2.34***	.20	10.35
Talkativeness	2.44***	.12	11.51
Affect	.68***	.11	1.98
Assertiveness	.93***	.06	2.54
Linguistic diversity	7.63***	.99	2054.02
σ^2	448.99		
τ_{00}	.58		
Deviance	300,072.60		
Chi-square (df)	448.31(15)***		

Note: The restricted maximum likelihood method is used for estimation. Predictor variables are estimates of the fixed effects, γs, with robust standard errors, and adjusted for overdispersion. N = 33,540. **p* <.05. ***p* <.01. ****p* <.001.

The Effects of Group Attributes on Conversation Creation

Finally, I hypothesized that several group attributes such as the size of the group (H4a), the amount of participation equality (H4b) and the group density or numbers of connections between members (H4c) would be positively associated with leadership, while the amount of turnover would be negatively associated with leadership (H4d). The means-as-outcomes model, which measures the group-level effect on the individual outcome variable is a good fit, χ^2 (11) = 71.80, p <.001, with a Deviance of 300,272.20. The group-level R^2 is .86, meaning that group attributes account for 86% of the variance in conversation creation. There is partial support for these hypotheses.

As shown in Table 4-12, larger discussion groups, measured in terms of the average numbers of participants every three months for a 20-month period, is positively related to conversation creation (B = 1.46, SE = .39, *exp(B)* = 4.30, t(11) = 3.71, p <.005). Discussion groups with more network density, or more connections between members in terms of message replies, are positively related to conversation creation (B = 1.46, SE = .21, *exp(B)* = 4.32, t(11) = 6.96, p <.001).

Contrary to prediction, discussion groups with more participation equality are not significantly related to conversation creation (p = .49). Nor are groups with a high member turnover related to conversation creation (p = .84).

Table 4-12. *Summary of Hierarchical Regression Analysis for Group Variables Predicting Leadership as Conversation Creation*

Variable	*B*	*SE*	*Exp(B)*
Intercept	2.30***	.11	9.97
Size	1.46**	.39	4.30
Participation Equality	–.54	.75	.58
Network Density	1.46***	.21	4.32
Turnover	.01	.05	1.01

Variable	*B*	*SE*	*Exp(B)*
σ^2	551.45		
τ_{00}	.10		
Deviance	306,928		
Chi-square (df)	71.80(11)**		

Note: The restricted maximum likelihood method is used for estimation. Predictor variables are estimates of the fixed effects, γs, with robust standard errors, and adjusted for overdispersion. N = 16. **p* <.05. ***p* <.01. ****p* <.001.

Summary

When we operationalize online leadership in terms of longer conversation (as opposed to a single reply), the results are the same. That is, posting and reply behavior, as well as tenure in the community, are positively related to conversation creation. Second, expansiveness and reciprocity are related to conversation creation, while brokering is not. Third, talkativeness, affect, assertiveness, and linguistic diversity are positively related to conversation creation. Again, reciprocity and linguistic diversity appear to have considerably larger effect sizes. Group size and density were also positively related to conversation creation, while participation equality and member turnover were not. The findings suggest that communication activity, social networks and language intensity and complexity not only contribute to receiving a reply to an online message, but actually creating a conversational thread or discussion around a topic or phenomenon.

Leadership as a Language Diffusion

Baseline Model and Intraclass Correlation

In the next analysis, the ICC is .02, which suggests that roughly 2% of the variance in leadership is between-groups and 98% of the variance is at the individual level. For the group-level predictors, the chi-square difference test is used, and the baseline model suggests that differences in discussion group size, equality of participation, network density and turnover rate are different than zero across groups (χ^2 (15) = 758.28, p <.001).

The Sociability and Trustworthiness of Conversation Creation

Again, the model is a good fit, χ^2 (15) = 623.79, p <.001, with a Deviance of 84861.34. The model also shows an R^2 of .97, suggesting that sociability and trustworthiness account for 97% of the within-subjects variance. Again, all three hypotheses are supported.

As shown in Table 4-13, the number of posts (H1a) is positively associated with language diffusion (B = .15, SE = .03, $exp(B)$ = 1.16, t(15) = 5.15, p <.001). The number of replies (H1b) is also positively associated with language diffusion (B = 1.41, SE = .21, $exp(B)$ = 4.12, t(15) = 6.87, p <.001). The tenure (H1c), or number of active days in the group, is also positively associated with language diffusion (B = .84, SE = .23, $exp(B)$ = 2.31, t(15) = 3.71, p =.002).

Table 4-13. *Summary of Random Coefficient Regression Analysis of Sociability and Trustworthiness Variables Predicting Leadership as a Language Diffusion*

Variable	*B (SE)*	*SE*	*Exp(B)*
Intercept	−1.99***	.09	.14
Number of Posts	.15***	.03	1.16
Number of Replies	1.41***	.21	4.12
Tenure	.84**	.23	2.31
σ^2	.13		
τ_{00}	.73		
Deviance	84,861.34		
Chi-Square (df)	623.79(15)***		

Note: The restricted maximum likelihood method is used for estimation. Predictor variables are estimates of the fixed effects, γs, with robust standard errors, and adjusted for overdispersion. N = 33,540. **p* <.05. ***p* <.01. ****p* <.001.

The Network Centrality of Language Diffusion

Again, I hypothesized that leaders are more likely to be expansive (H2a) and reciprocal (H2b) in their connections with others, and more likely to serve as brokers between otherwise disconnected participants (H2c). The model is a good fit, $\chi^2(15) = 629.50$, $p < .001$, with a Deviance of 93,910.40. This model also shows an R^2 of .96, meaning that the measures of centrality account for 96% of the variance in the amount of language diffusion.

As shown in Table 4-14, the first and second hypotheses are supported. Expansiveness, or outdegree centrality, is positively associated with language diffusion (*B* = 1.59, *SE* = .45, *exp(B)* = 4.92, *t*(15) = 3.58, *p* =.003). Reciprocity is positively associated with language diffusion (*B* = 2.01, *SE* = .31, *exp(B)* = 7.46, *t*(15) = 6.43, *p* <.001). Contrary to prediction, brokering (i.e., betweenness centrality) is not related to language diffusion (*p* = .09), although we see a possible trend.

Table 4-14. *Summary of Random Coefficient Regression Analysis of Social Network Variables Predicting Leadership as a Language Diffusion*

Variable	*B (SE)*	*SE*	*Exp(B)*
Intercept	−1.61***	.10	.20
Expansiveness	1.59**	.45	4.92
Reciprocity	2.01***	.31	7.46
Brokering	−.65	.36	.52
σ^2	.96		
τ_{00}	.15		
Deviance	93,910.40		
Chi-Square (df)	629.50(15)***		

Note: The restricted maximum likelihood method is used for estimation. Predictor variables are estimates of the fixed effects, γs, with robust standard errors, and adjusted for overdispersion. N = 33,540. **p* <.05. ***p* <.01. ****p* <.001.

The Language of Language Diffusion

Next, I predicted that leaders would be more likely to demonstrate higher frequencies of talkativeness (H3a), affect (H3b), assertiveness (H3c) and linguistic diversity (H3d). The model is a good fit, $\chi^2(15) = 542.91$, $p <.001$, with a Deviance of 193,782.24. This model also shows an R^2 of .28, meaning that these language variables account for 28% of the variance in the amount of language diffusion. All four hypotheses are supported.

As shown in

Table 4-15, talkativeness, measured in terms of the average message length, is positively associated with language diffusion ($B = 2.65$, $SE = .27$, $exp(B) = 14.08$, $t(15) = 9.95$, $p <.001$). Affect, measured in terms of the relative frequency of emotional words in a message, is positively related to language diffusion ($B = .76$, $SE =.15$, $exp(B) = 2.14$, $t(15) = 5.14$, $p <.001$). Assertiveness, measured in terms of the relative frequency of certainty words in a message, is positively related to language diffusion ($B = .94$, $SE =.05$, $exp(B) = 2.58$, $t(15) = 19.13$, $p <.001$). Finally, linguistic diversity, measured in terms of the type/token ratio, is positively related to language diffusion ($B = 7.57$, $SE =1.37$, $exp(B) = 1939.20$, $t(15) = 5.52$, $p <.001$).

Table 4-15. *Summary Of Random Coefficient Regression Analysis For Language Variables Predicting Leadership as Language Diffusion*

Variable	*B*	*SE*	*Exp(B)*
Intercept	–.75**	.20	.47
Talkativeness	2.64***	.27	14.08
Affect	.76***	.15	2.14
Assertiveness	.94***	.05	2.58
Linguistic diversity	7.57***	1.37	1939.20
σ^2	18.87		
τ_{00}	.54		
Deviance	193,782.24		
Chi-square (df)	542.91(15)***		

Note: The restricted maximum likelihood method is used for estimation. Predictor variables are estimates of the fixed effects, γs, with robust standard errors, and adjusted for overdispersion. N = 33,540. *p <.05. **p <.01. ***p <.001.

The Effects of Group Attributes on Language Diffusion

Finally, I hypothesized that several group attributes such as the size of the group (H4a), the amount of participation equality (H4b) and the group density or numbers of connections between members (H4c) would be positively associated with leadership, while the amount of turnover would be negatively associated with leadership (H4d). The means-as-outcomes model, which measures the group-level effect on the individual outcome variable, is a good fit, χ^2 (11) = 34.61, p <.001, with a Deviance of 204,628.40. The group-level R^2 is .92, meaning that group attributes account for 92% of the variance in language diffusion. There is partial support for these hypotheses.

As shown in Table 4-16, larger discussion groups, measured in terms of the average numbers of participants every three months for a 20-month period, are positively related to language diffusion (B = 1.34, SE = .26, $exp(B)$ = 3.81, t(11) = 5.09, p <.005). Discussion groups with more network density, or more connections between members in terms of message replies, are positively related to language diffusion (B = 1.54, SE = .26, $exp(B)$ = 4.67, t(11) = 8.84, p <.001).

Contrary to prediction, discussion groups with more participation equality are not significantly related to language diffusion (p = .99). Nor are groups with a high member turnover related to language diffusion (p = .43).

Table 4-16. *Summary of Hierarchical Regression Analysis for Group Variables Predicting Leadership as Language Diffusion*

Variable	*B*	*SE*	*Exp(B)*
Intercept	–.71***	.11	.49
Size	1.34**	.26	3.81
Participation Equality	–.01	.78	.99
Network Density	1.54***	.17	4.67

Variable	*B*	*SE*	*Exp(B)*
Turnover	–.02	.04	.97
σ^2	26.12		
τ_{00}	.05		
Deviance	204,628.40		
Chi-square (df)	34.61(11)**		

Note: The restricted maximum likelihood method is used for estimation. Predictor variables are estimates of the fixed effects, γs, with robust standard errors, and adjusted for overdispersion. N = 16. *p <.05. **p <.01. ***p <.001.

Summary

When online leadership is measured in terms of actually spreading word choices, the same independent variables are positively related to leadership. That is, posting and reply behavior, as well as tenure in the community, are positively related to conversation creation. Second, expansiveness and reciprocity are related to conversation creation, while brokering is not. Third, talkativeness, affect, assertiveness, and linguistic diversity is positively related to conversation creation. Again, reciprocity and linguistic diversity appear to have considerably larger effect sizes. Group size and density were also positively related to conversation creation, while participation equality and member turnover were not. The findings suggest that communication activity, social networks and language intensity/complexity not only contribute to receiving a reply to an online message, or creating a conversation, but actually spreading specific language use or ideas. Taken together, these findings provide strong support for a model of online leadership that includes both the quantity of participation and the quality of the messages that are contributed.

Summary of Results

Table 4-17 summarizes the various hypotheses tested in this section, along with a note on whether or not they were supported. In the next chapter, I discuss these findings along with their implications in light of previous research.

Table 4-17. *Summary of Hypotheses and Results*

Summary of Hypotheses	*Result*
H1. Sociability and the Posting Behavior of Leaders	
a. Posting Messages	Supported
b. Replying to other Messages	Supported
c. Longer Tenure	Supported
H2. Centrality and the Social Networks of Leaders	
a. Expansive connections	Supported
b. Reciprocity	Supported
c. Brokering	**Not Supported**
H3. The Language of Leaders	
a. Talkativeness	Supported

b. Affect	Supported
c. Assertiveness	Supported
d. Linguistic Diversity	Supported
H4. The Effect of Group Attributes on Leadership	
a. Group Size	Supported
b. Participation equality will be positively associated with leadership.	**Not Supported**
c. Density	Supported
d. Turnover	**Not Supported**

Chapter 5. Discussion

Summary

The purpose of this dissertation is to examine the communication behaviors of leaders in online discussion groups. I define online leaders as those participants who spark replies and conversation or influence the language and topical focus of their groups. I examine the extent to which certain linguistic and social network characteristics were associated with leaders, and whether attributes of the groups themselves, such as size, connectedness and stability, affect social influence.

In order to examine these issues, I rely on GOOGLE GROUPS, the current manifestation of *Usenet*, which allows users to post messages and reply to others, creating conversational threads or discussions. These discussion groups focus on a variety of topics such as breast cancer (i.e., health and support groups), gun rights (i.e., politics), open source software (i.e., science and technology) or quilting (i.e., hobbies and recreation). By focusing on all messages and user interactions that took place in sixteen randomly selected groups from June 2003 through January 2005, my sample comprises roughly 500,000 messages from 33,540 participants.

I also rely on a multi-method approach to examine these questions. First, I rely on user logs to capture the frequency of communication and time in the community. I also use social network analysis (Borgatti, et al., 1999) to identify the centrality and interaction patterns of all participants in each group, as well as the structural signatures of the groups themselves. Finally, I utilize a dictionary-based word-frequency analysis (Pennebaker, et al., 2006) to identify linguistic characteristics found in the messages of each participant.

Leadership is measured in several ways, all of which are extensions of the influence diffusion model (Matsumura, Ohsawa, & Ishizuka, 2002a). First, I measure a leader's popularity in terms of the number of replies he receives, what I call a reply trigger. I also measure the length of conversation that is spurred by each participant, which I refer to as conversation creation. Third, I trace the number of content-bearing words that cascade throughout these threads, which I refer to as language diffusion. The findings listed below were consistent across all three conceptualizations of online leadership.

Examining the post-and-reply behavior of leaders as well as their tenure in the community revealed that communication activity is indeed related to the ability to influence other members of the group. I also identify three key social interaction behaviors: (a) expansiveness, which is measured by the number of replies that a participant provides to others (i.e., outdegree centrality); (b) reciprocity, which is measured by the number of reciprocal links between two participants, in which they reply to each other; and (c) brokering, which is measured as the number of replies that a participant receives from two participants who do not connect to each other (i.e., betweenness centrality). Both expansiveness and reciprocity increase social influence, but brokering does not have an effect.

I also identify the most prevalent linguistic traits associated with social influence. These include: (a) talkativeness, which is measured in terms of the average length of messages; (b) affect, or language with a positive or negative emotional valence; (c) assertiveness, which exudes certainty, confidence or resoluteness; and (d) linguistic diversity, which represents lexical complexity or vocabulary richness. All four linguistic variables increase social influence.

Finally, I argue that size, connectedness, equality and turnover are factors that can influence the ability for leaders to influence others. These four variables are conceptualized as: (a) group size, which is calculated as the average number of authors who contributed to the group every three months during a 20-month period; (b) participation equality, which measures complexity in terms of the proportion of participation from all members of the group; (c) network density, which stands for the proportion of linkages between members of the group to the potential number of linkages; and (d) group stability, which is captured by the percentage change in membership every three months for a 20-month period. Only size and density were positively related to online leadership.

Again, the results are the same across all three conceptualizations of online leadership. Whether social influence in CMC settings is understood in terms of triggering a reply, sparking a longer conversation or spreading actual words and ideas across a discussion thread, the findings show that a specific set of communication features, including specific linguistic devices, increase an individual's ability to influence other members of an online community. The findings also show that some group characteristics can help facilitate this ability. In this final chapter, I summarize the previous data analyses and discuss the findings in light of research in CMC.

Communication Activity, First and Foremost

The findings show that sheer communication activity is central to being influential. This is because one cannot influence others with zero contribution. It has to start somewhere, and individuals who post are more likely to be leaders in the community. This research also shows that the more individuals contribute to the group, the higher the chance that they will start a conversation or spread an idea. This is important because in most online communities the bulk of information is produced by a subset of users. For example, Smith (1999) finds that 18,000 people contribute 67,000 messages to Usenet daily (p. 209). WIKIPEDIA, widely regarded as a successful community because of its population and amount of contribution (on the order of 2 million registered users), relies on a small group of editors to do the majority of work (Kittur & Kraut, 2008). The small subset of online contributors is also evident by the substantial lurker populations in most online communities, who enjoy perusing content without adding to it (Nonnecke, Andrews, & Preece, 2006).

Yet contribution is integral to the success of a community. If no one replied to a message, the community would have no purpose. What is 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, my findings show that posting behavior is outweighed by reply behavior, or responding to the posts of others (more specifically, the coefficients for reply behavior show a much greater magnitude and odds ratio). There are a couple explanations for this. Replying to others is single-handedly the strongest motivator for getting users to come back or become committed to a community (Joyce & Kraut, 2006). By replying to others, users get a chance to share a new idea, or build upon existing ones. To other members of the community, a response shows that the message is worth reading, an appeal that should increase as the thread lengthens, giving the impression of a hot topic. Responding to others increases the opportunity to get threads started, or control the focus of the topic. The side effect of this is also positive: it gives the original poster more incentive to contribute and stay in the community because it increases a sense of belonging.

The importance of credibility, as outlined by social influence theorists (O'Keefe, 2002), also plays an interesting role online. While the findings show that leaders are more likely to have longer tenure in the community, it may not always be clear how long users have been active. For example, in some message-board communities, registered users often have an associated name, avatar and status message (e.g., 'super user', or 'posted 1,000 messages'), which is a clear indicator of tenure. The discussion groups used here do not have those distinguishing features, but they do include actual email addresses. The issue here is that newcomers to a community cannot easily identify the veterans of the community unless they spend time watching the conversation flow or search through post archives. However, veterans are likely to know the tenure of frequent posters based on experience. So it might be the case that veterans engage with other veterans based on tenure. Even so, the findings still support the argument that reputation is an important aspect of online trust and transactions (Resnick, Kuwabara, Zeckhauser, & Friedman, 2000), even if those transactions are communication.

In sum, posting and replying to messages, along with tenure in the community, are strong predictors of online leadership. This is partly explained because message activity is integral to keeping the community thriving, yet replying to messages actually increases influence, and perhaps more so than simply posting. I will discuss the importance of this more in the next section on the social network behaviors of leaders.

The Importance of Building Relationships

In addition to replying to others more often, 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 members can come and go without a second thought.

Wellman & Gulia (1999) argue that online community interactions go beyond simple information exchange which would facilitate "narrow, specialized" relations—and, instead, encourage supportive, loyal relationships (p. 4). My findings support this notion: leaders engage in reciprocal links by replying to those who engage them. This suggests that leaders are often 'reaching out' to the community and that these networking behaviors increase their ability to spark a conversation or spread an idea. Equally important, the findings support Walther's (1996) argument that online technologies are interpersonal in nature, and extend it by showing that interpersonal behavior has an effect on communication outcomes such as personal influence.

The findings also provide more insight into previous research examining who really does the work in online communities (Butler, Sproull, Kiesler, & Kraut, 2002). The results suggest that when people are close to others (even in terms of sharing communication messages), they are more likely to adopt an idea (Granovetter, 1978) and that information often flows through networks in a cascade (Watts, 2002). Both of these ideas resonate with reciprocity; as individuals interact back-and-forth, they are more likely to be influenced by one another, and in turn spread information to other parts of the network.

More broadly, the findings lend support to previous work suggesting that diffusion is greater when people are associated based on common interests (Huberman & Adamic, 2003). Perhaps the ways in which online groups are often organized — by a shared topic (e.g., Philip Seymour Hoffman fan club) or novelty (e.g., TWITTER reaches a critical mass as it becomes a media darling), invites bonding behavior and reciprocity. Online communities often resemble Putnam's (2000) example of bonding social capital (i.e., a country club), and are increasingly niche. For example, social networking services like MYSPACE and FACEBOOK have millions and millions of users, yet there are niche versions of social networks based on ethnicity (e.g., African-American culture), specific interests or hobbies (e.g., bird-watching or golfing), or subculture (e.g., vampires). These niches likely increase the bonding behavior of the community, explaining why reciprocity is such an important feature of the communication networks.

In fact, the high correlation between some of the centrality measures suggests that across the broader community, everyone is connecting to others, receiving connections and connecting those that connect to them. This is somewhat surprising in a voluntary discussion group, where people can come and go at will and attrition is rampant. Similarly, one would think that inhabiting a group of roughly 360 authors on average would make it difficult to connect to everyone. Yet, all participants are repeatedly connecting to the same participants, suggesting that computer-mediated contexts such as online communities facilitate connectivity and participation, perhaps more so than real-world organizations (Sproull & Kiesler, 1991).

Contrary to expectation, brokering (i.e., betweenness centrality) was not a significant predictor of online social influence. While there has been strong support that serving as a broker and regulating the flow of information between groups (in effect, filling 'structural holes') has distinct competitive advantages in an organizational setting (Burt, 1992), and that leaders often rely on structural holes to accumulate social capital (Burt, 1999), it may not be as prevalent or important in large-scale online communities. According to these findings, everyone in the community can serve as a broker, and leaders are not exceptional in this way. One explanation is that all information in online discussion groups is transparent. And because people can read as many posts as they want, without requiring them to respond, brokering does not hold the same power. Second, these networks are generally sparse, meaning that there are many cases where participants do not link. These could result in most participants having high betweenness scores, which is evident in the high average I found.

Telling an Interesting, Passionate or Compelling Story

The results show that talkativeness, affect, assertiveness and linguistic diversity are linguistic traits positively associated with online leaders. On the one hand, this might represent a set of personality traits manifesting in CMC (Gill, 2004), which could serve as a feature of influence (Cialdini, 1994); on the other hand, these features might serve to increase the salience of the messages by creating interesting, passionate or compelling stories.

First, the frequency of communication activity shows up as an important factor again. The finding that online leaders tend to produce lengthier messages is interesting here because previous research has suggested that CMC 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 technical or sociocultural constraints, leaders still engage in lengthier dialogues or monologues.

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 probably has more to 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), which could be in play here. Leaders may be engaging in conversations using positive and negative emotional language during deliberations to make a point. This passion may enhance argument because it adds credence or confidence in the idea being shared.

However, leaders might also be engaging in more complex social behaviors by utilizing emotional intelligence, which refers to the ways in which people express feelings to present themselves to others (Salovey, 1990). In fact, emotional language may be useful in developing rapport with another individual (Barry & Oliver, 1996), which supports the previous findings on building relationships with other group members. It could also be useful in maintaining a positive attitude or mood within the group (Bono & Ilies, 2006). For example, a newcomer might visit a group, see that it is generally pleasant, and feel more comfortable in participating. By contrast, a negatively charged group might deter some members, while encouraging others to join the fray.

The assertiveness finding suggests that online leaders exude more confidence or certainty in their online communication, and that powerful language, like affect, seems a successful persuasive device that leaders seem to utilize in this forum when trying to get a point across or debate an issue. And scholars have argued that powerful language is something generally prevalent in discussion groups. Herring shows how aggressive behaviors prevail online, although not necessary to the benefit of all participants (Herring, 2003). The extreme version of this is “flaming”, a hostile interaction between users often involving profanity, obscenity or insults, which some scholars relate to levels of assertiveness (Alonzo & Aiken, 2004, p. 205).

Based on the magnitude of the coefficients and the event ratios[9], linguistic diversity has the strongest influence on 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/readers have been shown to relate lexical diversity to a speaker's/writer's status, intellectual competence and communication ability (Ng & Bradac, 1993). As Ng & Bradac (1993) describe, "…lexical diversity is an attributionally rich feature of language information about communicator affiliation, traits, and states" (p. 44), which links directly to social influence theories on 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.

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.

---

[9] In Poisson regression, the coefficients must be converted using an antilog, or *exp(B)*. This is commonly referred to as an event ratio, but it's similar to the concept of an odds ratio used in logistic regression.

The Group Effect

Some group characteristics do have an effect on online leadership. First, group size has a significant, positive association to leadership, This is interesting because in business environments, the size of an organization can limit the connections between members, creating more social distance and hinder the emergence of leaders (Bass, 1990). Even in small groups, as the size increases, participants have fewer opportunities to talk or engage in leadership activities (Hare, 1976). Furthermore, group size tends to affect normative behaviors, such as inhibiting conduct that stands out for fear of rejection (Brown, 2006).

However, group size has an opposite effect in online communities. More members means more opportunities to influence and spread ideas. It usually means more communication activity, although in some groups size does not produce this effect . And studies of CMC also suggest that there is often more participation and longer interactions when compared to face-to-face environments (Bordia, 1997), and that participants have less normative pressure and show less inhibition (Suler, 2004). As Castells (1996) points out, online networks allow information to be produced and exchanged at a low cost. This should make it easier for participants to communicate more often and ultimately emerge as leaders.

I measured the complexity of the organizations in terms of participation equality, which included the proportion of message posts that all members of the group made, as well as network density, which essentially represents how *connected* the group is. Participation equality is not a significant predictor of online leadership, suggesting that leaders emerge at the same rate regardless of whether everyone in the group participates equally. Again, this could be due to the nature of CMC. Early studies demonstrate more equal participation among the members of a group (Kiesler & Sproull, 1992) and less domination by a few individuals (McLeod, 1992). In addition, my findings suggest that groups have low participation equality on average.

Network density, on the other hand is a significant predictor of individual social influence, suggesting that more connections between members of the group are positively associated with leadership. This might be explained by the centrality measures, in which popularity, expansiveness and reciprocity are all connected to leadership. And while participation equality includes posting behavior, which may never receive a reply, network density is measured by the number of replies that a participant receives. This makes the connections between people more clear. Some suggest that the scale-free network structure of the internet creates major communication hubs connected by many smaller hubs, resulting in close connections between all participants (Barabási & Crandall, 2003). This would allow leaders to communicate more and more as the network connections grow. This finding also shows that highly connected groups represent a level of social support, i.e. actual relationships between members, that results in longer conversations and more idea sharing among multiple authors. Therefore, even large-scale online networks exhibit relationship building and supportive behaviors that many CMC scholars have previously pointed out (Baym, 2000; Cummings, Butler, & Kraut, 2002; Hampton & Wellman, 1999; Rheingold, 2000; Wellman & Gulia, 1999).

Group turnover, which is measured in terms of percentage change in membership, was not found to be significantly associated with online leaders. While groups in this study maintain a *status quo* on average, there is a dynamic nature of online groups involving serious attrition. Studies show that whether the community involves discussion groups, open source software projects, or even online gaming communities, at least a quarter of participants leave quickly (Ren, Harper, Drenner, Kiesler, Terveen, & Riedl, under review). Such is the case in this study: only fifty or so authors of over 30,000 contributed through the entire period examined. However, many authors participate for months at a time and serve as the foundation for interaction and social influence regardless of whether newcomers arrive or some veterans depart. And although commitment to the group can lead to its success, the findings here show that high member turnover does not negatively affect the ability of a leader to influence those who remain.

In conclusion, the results have provided a deeper insight into the social behavior of online leaders, including their communication frequency and interaction patterns, as well as the specific linguistic traits that increase influence. Overall the findings suggest that communication activity remains a prerequisite for being influential; however, building relationships, bonding with others, and creating interesting, passionate or compelling messages distinguish online leaders from their counterparts. In addition, while size and connectivity among members increases the ability to influence others, the equality and turnover of such groups do not.

Implications

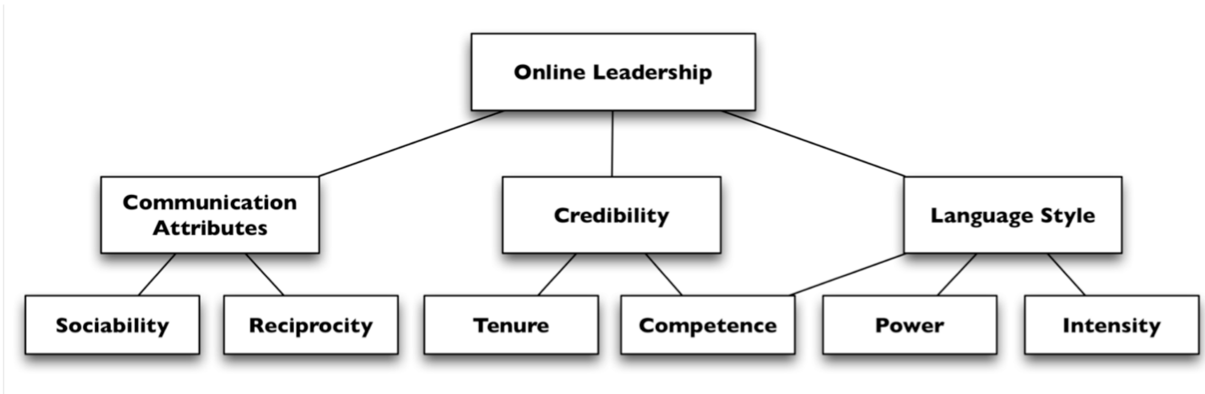
A New Framework of Online Leadership

The central theoretical contribution of this project is to provide a comprehensive conceptualization of online leadership. Previous work examining online leaders has focused on the babble hypothesis, which states that the sheer quantity of speech results in the emergence of leaders in otherwise leaderless groups (Misiolek & Heckman, 2005; Yoo & Alavi, 2004). Leadership scholars have challenged this notion, arguing that quality must be more important than quantity (Bass, 1990), yet no one has provided an empirically-validated set of communication features to represent leadership. This research serves as a foundation for developing a theory of online leadership, in which communication activity is coupled with relationship-building behaviors and specific linguistic devices.

Figure 5-1 illustrates a framework for understanding online leadership based on the results of this research. It includes sociability and reciprocity, both general communication-activity attributes. Sociability refers to communication in much the same way as the babble hypothesis and represents the frequency of incoming and outgoing communication. Centrality refers to interaction processes, including relationship-building or bonding behavior.

The credibility of a person is critical to his being perceived as a leader, which, hinges on being recognized as a long-time member of the community (i.e., tenure, or being a veteran) and being competent, which manifests itself in terms of expertise and knowledge, or high-quality communication skills (i.e., good writing, wit, eloquence). Finally, the language style itself plays a part. Power (i.e., confidence and assertiveness) and intensity (i.e., emotional or affective processes) are positively related to online leadership.

Figure 5-1. A framework for online leadership.



TWITTER provides a specific example. My framework suggests that people are more likely to have many followers if they are active and engage other members of the community by responding or integrating them into their feeds (known as @replies or re-tweeting). They are also likely to have been a part of the community for a long-time (possibly early adopters) as evidenced by their many updates and date/time-stamps. They must appear competent and credible—that they *know* what they are talking about, as evidenced by their bios, writing ability and topic choice. Most importantly, leaders must be interesting, humorous, passionate and compelling in their actual messages; otherwise, they are likely to be dropped. Each of these features works to enhance a leader's ability to influence other members, but they are not necessarily dependent on each other. However, a leader's ability to influence is maximized when all features work together.

Methodological Implications

This research utilizes a multi-method approach that is indicative of interdisciplinary research, and shows that combining analytical tools from different disciplines offers many insights into the field of computer-mediated communication. Specifically, I demonstrate the utility of word frequency analysis to analyze human language in online settings and social networks to develop a comprehensive understanding of relationship development in online groups.

These techniques also point toward the importance of quantitative analysis in the study of computer-mediated communication. While qualitative methods such as ethnography or interview data are important tools for discovering user attitudes and perceptions, quantitative approaches allow researchers to examine natural social behaviors without intrusion. Certainly, the combination of qualitative and quantitative methods offers the most robust understanding of internet user behavior, but even, log data analyses alone can tell us much about communication trends and social activities of online users.

Second, this dissertation has highlighted the importance of considering aspects of the group in the analysis of online interaction. Understanding the influence of the context that an individual inhabits — the *dynamics* of the group — adds additional insight into why a user might behave in a particular way. This is not just a concern for discussion groups, which number in the millions and vary dramatically in topic areas. It is also important in the blogosphere, in chat rooms, or within virtual work groups, all of which vary in size and focus. In fact, group-level variables can also be included when comparing across several similar web applications, such as users with accounts on FACEBOOK, MYSPACE and

FRIENDSTER. Differences in population, activity, content, or even features will influence user behavior, and researchers need a technique such as HLM in order to examine these nuances.

Practical Applications

This dissertation also has several practical applications for business and, education as well as for the design of online communities and other web applications that foster social interaction. The business application concerns marketers wanting to identify what the hottest topics are on the web, and who produces and disseminates them . Marketers want to know: Who are the most influential bloggers? Who spreads news stories or mobilizes users through FACEBOOK and TWITTER? As an advertiser, whom do I want to reach first? My dissertation provides a framework for identifying the most influential members of a community to facilitate targeting them for important initiatives.

Targeting influential members is also useful for learning communities, especially since they tend to build relationships with their peers. Integrating discussion forums and blogs into everyday learning environments should increase the speed with which information is delivered and consumed. And leaders of these groups are more likely to influence others regarding knowledge, which can be an excellent extension to the teacher.

Finally, this dissertation research is useful for technology designers, and important for community managers or group moderators. They can encourage or reward the behaviors of those who are targeted as emerging leaders. Take, for example, a virtual work group where no leader has been assigned. An analysis of the interactions within the group over a set period time of time would reveal which workers emerge as leaders, which gives management insight into the talent pool and employees a chance for promotion. Likewise, in an e-learning application, educators can use these models to identify which students are leading their learning groups. Because peer learning has been widely considered to be a powerful tool in education (Blumenfeld, Marx, Soloway, & Krajcik, 1996), targeting student leaders early on can help educators to better construct appropriate learning groups. More generally, automatically detecting and recognizing leadership often leads to more participation (Chan, Bhandar, Oh, & Chan, 2004; Hummel, Burgos, Tattersall, Brouns, Kurvers, & Koper, 2005; Ling, Beenen, Ludford, Wang, Chang, Li, Cosley, Frankowski, Terveen, Rashid, Resnick, & Kraut, 2005).

A second implication for design is developing better recommendation systems. If new users join a community, they might be recommended to a local leader, who can welcome them, fostering a sense of belonging, which some scholars have shown to be associated with greater participation (Hemetsberger, 2001). Indirectly, these recommendations would serve as a reputation system, which helps users evaluate content integrity and user status (Baker, Jensen, & Murphy, 1988). For leaders, feeling that they have a good reputation impacts their willingness to help others, especially over time (Drago & Garvey, 1998)

Recommendation systems are especially useful when interests or content contribution are connected with users. For example, a system can identify the various topic areas based on the language of leaders, mark those as expertise areas, and recommend other users to these experts. Or, given that leaders in a community are influential, systems might be able to detect which products or ideas could be most attractive to other users. At a higher level, matching users based on their participation rates, interaction patterns and language could result in ideal recommendations for a friendship or a possible romantic relationship.

Limitations

Although this research has several strengths, there are also limitations. First, there is no way to tell if other members of the community perceive these participants as leaders. This does not affect the results of this dissertation, or the evidence that users who contribute the most influence conversation or the diffusion of ideas, but it is still important to know if other users see these participants as the active leaders of the group. Intuitively, users who contribute the most should be the most salient members of the group, and previous findings on emergent leaders in virtual groups suggest that these are the members most likely to be perceived as leaders (Misiolek & Heckman, 2005; Sarker & Grewel, 2002; Yoo & Alavi, 2004). However, additional qualitative (in which users are interviewed) or quantitative analysis (in which users are surveyed) would confirm this suspicion.

Second, trolls and spammers may still penetrate these groups, causing some noise in the data. While several steps were taken to identify spammers and binary posters (i.e., those simply posting images or files), and some were removed from the data set, there is still a chance that spammers are included in the set. However, most spammers come from a small set of distinguishable IP addresses, which can easily be detected and disqualified

(Ramachandran & Feamster, 2006). One of the features of GOOGLE GROUPS is the ability to mark a message as spam, making it easy for users to moderate and regulate each discussion group. Together, this suggests spammers can be detected and should not pose a major issue here.

Trolls, on the other hand, who watch discussion groups and post messages to be purposely provocative, often relying on emotionally-laden messages (Donath, 1998), could be affecting some of the data. This is especially problematic in non-mainstream forums (e.g., feminism), but trolls can be regulated by ignoring or blocking their messages (Herring, et al., 2002). While it's not clear if trolls were regulated in this community, GOOGLE GROUPS does offer users a way to block messages or report them as spam. That said, it would seem unlikely that trolls would be able to reach the levels of contribution necessary to be considered leaders. If anything, their messages diminish the magnitude of some of the linguistic characteristics (e.g., affective or emotional language, or certainty and confident language) or the network variables (e.g., trolls might receive more incoming links of people arguing with them or asking them to leave). However, removing trolls from the data would not be appropriate, as they represent a phenomenon that has always been part of internet culture (Tepper, 1997).

It is also important to examine the ability to draw generalizations from of these findings. *Usenet* represents 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" (p. 30). Tepper (1997) argues that different groups attract distinctive readers and posters from the larger heterogeneous *Usenet*, and adherence to group norms and standards lead to unique subcultures. Although the exchange of information is a primary function of *Usenet*, it allows for much identity exploration and deception, including the infiltration of trolls, spammers and name-switching (Donath, 1998) because anyone with an email address can join and groups are basically unmoderated (Tepper, 1997). This combination of group norms, emerging subcultures and identity freedom allows for unique online interactions that may not be evident in other forms of online communities.

What *Usenet* does share with many online applications is basic communication processes. Posting content and receiving feedback from others serves as the foundation of many internet activities. What is the essence of email or an online chat, but the very basic post-and-response? Social network services flourish when users post updates, pictures and profile information, and others validate them. Photo- and video-sharing web sites and blogs alike thrive on user comments to posted content in order to spark discussion or controversy. Message boards, Q&A sites, fan sites and the like are really an evolved state of *Usenet*, using the same threaded conversation structure, but including additional features such as adding a profile and other aspects of identity (e.g., a handle, an avatar, user status, etc.) and more stylistic message production (e.g., cool fonts, colors, multimedia, etc.). In short, *Usenet* authors represent a unique set of internet users, but they do engage in the most common internet communication patterns.

The second concern with the generalizability of *Usenet* involves its age. One could argue that the users who first engaged in Usenet in the 1980s might represent a unique population. They could be portrayed as early adopters or as 'cyber nerds' focused on cultural artifacts that lay on the fringe of society. However, these interactions took place in 2003 – 2005, when the internet was common in most homes (at least in the United States) (Rainie & Horrigan, 2004), and participating in online communities was not a very unusual activity (Horrigan, 2001). Although *Usenet* is a distinctive type of online community, it is possible that a more general internet audience had begun using GOOGLE GROUPS by 2003.

Future Directions

Online discussion groups represent a distinct set of features that distinguishes them from other online communities and exhibits their unique cultural practices. Therefore, future research in this area should analyze how language and social networks increase social influence in other online communities such as blogs, social network services, content-sharing web sites, chat rooms or other kinds of message boards (e.g., PROBOARDS or PHPBB, which offer free online forums similar to GOOGLE GROUPS). While each of these online communities represents a unique set of technical features or user culture, examining leadership within them will offer a more robust understanding of user behavior, which can help strengthen a general theory of online leadership and social influence.

I also hope to tease apart cultural practices in different online communities. For example, to what extent does leadership and social influence differ in open source software communities, support groups or political activist groups? Does the technological platform (such as a mobile device versus a desktop computer) impact social influence? Are there aspects of social interaction that are shared across all groups or technological platforms? Understanding how technological affordances and cultural values influence organizational patterns provides insights into the design of Web 2.0 services, intelligent agent-based systems and applications that facilitate communication and collaboration in institutional settings such as higher education, government and business.

Second, I intend to focus on the dynamic nature of leadership emergence and social influence. For example, how long does it take to become a leader in a group? Once a leader emerges, how does that impact the social dynamics of the rest of the group? Do online communities require a critical mass to succeed, or can a small group of influential members foster growth? Understanding changes over time advances our knowledge of group dynamics and information diffusion, and the outcomes of this research will allow us to forecast how technological interventions impact social behavior and group performance.

Finally, future work could refine the analytical tools used in this dissertation. For example, language diffusion was measured in terms of shared words or phrases, but it does not distinguish what types of words are being mirrored. These could be topical (i.e., mirroring persons, places or things) or they could represent sentiment (i.e., agreeing that something is bad or good). This would inform whether social influence manifests more in terms of a person or product, or in terms of an opinion or feeling.

Likewise, other measures such information entropy or vector-based approaches, which have also been shown to be applicable to online community language use, might be able to at least confirm the findings using term frequency evaluation (Huffaker, Jorgensen, Iacobelli, Tepper, & Cassell, 2006). Similarly, the dictionaries used for the linguistic traits could be expanded to include more words, once they were validated as accurate measures of a particular psychosocial dimension. Likewise, automated text analysis and computational linguistics can be used to uncover how members of online communities identify and rate leaders through name-dropping and citations to previous contributions. These extrinsic dimensions of leadership help us understand the impact of credibility and expertise in online communities, and can better inform the design of reputation and recommendation systems.

Additionally, the social network analysis used here, which focuses primarily on measures of centrality, could utilize some of the more complex studies of network behaviors, such as sub-groups and cliques (Wasserman & Faust, 1994) or even the advanced components found in exponential random graph modeling (e.g., 2-stars, K-Star, triangles) (Snijders, Pattison, Robins, & Handcock, 2006). These advanced techniques might provide more insight into the dynamics of these large networks. In short, extending the breadth and depth of each measure used in this dissertation offers an important future direction for understanding online leadership, the dynamics of complex networks and the diffusion of information.

Conclusion

In conclusion, this dissertation provides a rich understanding of what constitutes an online leader. By examining communication activity, social interaction and language use in large-scale online discussion groups, it identifies a set of traits that increase an online leader's ability to set agendas or frame discussions. Specifically, it shows that frequent posting and replying to messages, message length and tenure in the community increase one's ability to trigger a reply, spark a conversation and spread an idea or word choice. Likewise, expansiveness and reciprocal social network behaviors, which are aspects of collecting and spending social capital, is positively related to online leaders. Finally, specific linguistic qualities such as affect, assertiveness and linguistic diversity, which are aspects of language intensity and complexity, are positively related to online leaders.

At the same time, this dissertation shows the importance of considering group attributes in the study of online behavior. It also shows that group size and connectedness positively influence online leaders and, therefore, the flow of information. Together, this dissertation presents a model of online leadership and demonstrates the utility of combining quantitative analytic tools such as text analysis and social network analysis, along with statistical evaluation that takes into account group-level variance. In effect, this dissertation provides a strong foundation for understanding the ways in which online leaders communicate, interact and influence those around them.

References

Adkins, M., & Brashers, D. E. (1995). The power of language in computer-mediated groups. *Management Communication Quarterly, 8*(3), 289-322.

Alavi, M. (1994). Computer-mediated collaborative learning: An empirical evaluation. *MIS Quarterly, 18*(2), 159-174.

Alonzo, M., & Aiken, M. (2004). Flaming in electronic communication. *Decision Support Systems, 36*(3), 205-213.

Andrews, D., Nonnecke, B., & Preece, J. (2003). Electronic survey methodology: A case study in reaching hard-to-involve internet users. *International Journal of Human-Computer Interaction, 16*(2), 185-210.

Aral, S., Brynjolfsson, E., & Van Alstyne, M. W. (2007). *Productivity effects of information diffusion in networks*: Social Science Research Network (SSRN).

Atkinson, A. B. (1970). On the measurement of inequality. *Journal of Economic Theory, 2*(3), 244-263.

Baker, G., Jensen, M., & Murphy, K. (1988). Compensation and incentives: Practice vs. theory. *Journal of Finance, 43*, 593-616.

Balkundi, P., & Kilduff, M. (2005). The ties that lead: A social network approach to leadership. *The Leadership Quarterly, 16*(6), 941-961.

Barabási, A. L., & Crandall, R. E. (2003). Linked: The new science of networks. *American Journal of Physics, 71*, 409.

Barry, B., & Oliver, R. L. (1996). Affect in dyadic negotiation. *Organizational Behavior and Human Decision Processes, 67*(2), 127-143.

Bass, B. M. (1990). *Bass & Stogdill's handbook of leadership: Theory, research, & managerial applications* (Third ed.). New York: The Free Press.

Baym, N. K. (1995). From practice to culture on usenet. In S. L. Star (Ed.), *The Cultures of Computing* (pp. 29-52). Cambridge, MA: Blackwell Publishers.

Baym, N. K. (2000). *Tune in, log on: Soaps, fandom, and online community*. New York: Sage Publications.

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

Bikhchandani, S., Hirshleifer, D., & Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy, 100*(5), 992-1026.

Bless, H., Bohner, G., Schwarz, N., & Strack, F. (1990). Mood and persuasion: A cognitive response analysis. *Personality and Social Psychology Bulletin, 16*(2), 331-345.

Block, L. G., & Keller, P. A. (1995). When to Accentuate the Negative: The Effects of Perceived Efficacy and Message Framing on Intentions to Perform a Health-Related Behavior. *Journal of Marketing Research, 32*, 192-203.

Blumenfeld, P. C., Marx, R. W., Soloway, E., & Krajcik, J. (1996). Learning With Peers: From Small Group Cooperation to Collaborative Communities. *Educational Researcher, 25*(8), 37-39.

Bono, J. E., & Ilies, R. (2006). Charisma, positive emotions and mood contagion. *The Leadership Quarterly, 17*(4), 317-334.

Bordia, P. (1997). Face-to-face versus computer-mediated communication: A synthesis of the experimental literature. *The Journal of Business Communication, 34*(1), 99-121.

Borgatti, S. P., Everett, M. G., & Freeman, L. (1999). UCINET V. Software for social network analysis. Natick, MA: Analytic Technology.

Bourdieu, P. (1991). *Language and symbolic power* (G. Raymond & M. Adamson, Trans.). Cambridge: Polity Press.

Bradac, J. J., Bowers, J. W., & Courtright, J. A. (1979). Three language variables in communication: Intensity, immediacy and diversity. *Human Communication Research, 5*(3), 257-269.

Bradac, J. J., Konsky, C. W., & Davies, R. A. (1976). Two studies of the effects of linguistic diversity upon judgments of communicator attributes and message effectiveness. *Speech Monographs, 43*(1), 70-79.

Brass, D. J. (1984). Being in the right place: A structural analysis of individual influence in an organization. *Administrative Science Quarterly, 29*(4), 518-539.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(6), 1482-1493.

Brief, A. P., & Weiss, H. M. (2002). Organizational behavior: Affect in the workplace. *Annual Review of Psychology, 53*(1), 279.

Briggs, P., Burford, B., De Angeli, A., & Lynch, P. (2002). Trust in Online Advice. *Social Science Computer Review, 20*(3), 321-332.

Brown, C. (2006). *Social psychology*. London: Sage Publications Ltd.

Brown, G., & Yule, G. (1983). *Discourse analysis*. Cambridge: Cambridge University Press.

Burgoon, M., Jones, S. B., & Stewart, D. (1975). Toward a message-centered theory of persuasion: Three empirical investigations of language intensity. *Human Communication Research, 1*(3), 240-256.

Burrell, N. A., & Koper, R. J. (1998). The efficacy of powerful/powerless language on attitudes and source credibility. In M. Allen & R. W. Preiss (Eds.), *Persuasion: Advances Through Meta-analysis* (pp. 203-216). Cresskill, New Jersey: Hampton Press, Inc.

Burt, R. S. (1992). *Structural holes: The social structure of competition*. Cambridge: Harvard University Press.

Burt, R. S. (1999). The social capital of opinion leaders. *The Annuals of the American Academy of Political and Social Science, 566*(1), 37-54.

Burt, R. S. (2000). The network structure of social capital. In R. I. Sutton & B. M. Staw (Eds.), *Research in Organizational Behavior* (Vol. 22, pp. 345-423). New York: El Sevier Science.

Butler, B. S., Sproull, L., Kiesler, S., & Kraut, R. (2002). Community building in online communities: Who does the work and why? In S. Weisband & L. Atwater (Eds.), *Leadership at a Distance: Research in Technologically-Supported Work* (Vol. 171-194). Mahwah, NJ: Erlbaum.

Cacioppo, J. T., & Petty, R. E. (1982). Language variables, attitudes and persuasion. In E. B. Ryan & H. Giles (Eds.), *Attitudes Towards Language Variation* (pp. 189-207). London: Edward Arnold.

Cacioppo, J. T., & Petty, R. E. (1989). Effects of message repetition on argument processing, recall, and persuasion. *Basic and Applied Social Psychology, 10*(1), 3-12.

Cameron, A. C., & Trivedi, P. K. (1998). *Regression analysis of count data*: Cambridge University Press.

Cameron, D. (1998). Performing gender identity: Young men's talk and construction of heterosexual masculinity. In J. Coates (Ed.), *Language and Gender: A Reader* (pp. 270-284). Malden, MA: Blackwell Publishers Ltd.

Cassell, J., Huffaker, D., Tversky, D., & Ferriman, K. (2006). The language of online leadership: Gender and youth engagement on the Internet. *Developmental Psychology, 42*(3), 436-449.

Castells, M. (1996). *The rise of the network society* (Second ed. Vol. 1). Malden, MA: Blackwell Publishing.

Chan, C. M. L., Bhandar, M., Oh, L.-B., & Chan, H.-C. (2004). *Recognition and participation in a virtual community.* Paper presented at the HICSS'04: 37th Annual Hawaii International Conference on System Sciences.

Cialdini, R. B. (1994). *Influence: The psychology of persuasion*. New York: William Morrow.

Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology, 94*(S1), 95.

Conger, J. A. (1991). Inspiring others: The language of leadership. *Academy of Management Executive, 5*(1), 31-45.

Corman, S. R., Kuhn, T., McPhee, R. D., & Dooley, K. J. (2002). Studying complex discursive systems: Centering resonance analysis of communication. *Human Communication Research, 28*(2), 157-206.

Crystal, D. (2001). *Language and the internet*. New York: Cambridge University Press.

Cummings, J., Butler, B., & Kraut, R. (2002). The quality of online social relationships. *Communications of the ACM, 45*(7), 103-108.

Cummings, J. N., & Cross, R. (2003). Structural properties of work groups and their consequences for performance. *Social Networks, 25*(3), 197-210.

Danet, B., & Herring, S. C. (Eds.). (2007). *The multilingual Internet: Language, culture and communication online*. Oxford: Oxford University Press.

Dawkins, R. (2006). *The selfish gene* (30th Anniversary ed.). New York: Oxford University Press.

Dearing, J. W., & Rogers, E. (1997). *Agenda-Setting*. Thousand Oaks, CA: Sage Publications.

Degenne, A., & Forse, M. (1999). *Introducing social networks*. London: SAGE Publications.

Donath, J. S. (1998). Identity and deception in the virtual community. In P. Kollock & M. Smith (Eds.), *Communities in Cyberspace*. London: Routledge.

Drago, R., & Garvey, G. (1998). Incentives for helping on the job: Theory and evidence. *Journal of Labor Economics, 16*, 1-25.

Eagly, A., & Chaiken, S. (1993). *The psychology of attitudes*. Fort Worth: Harcourt Brace Jovanovich College Publishers.

Eckert, P. (1989). *Jocks & burnouts: Social categories and identity in the high school*. New York: Teachers College Press.

Eysenck, H. (1967). *The biological basis of personality*. Edison, NJ: Transaction Publishers.

Finholt, T., & Sproull, L. S. (1990). Electronic groups at work. *Organization Science, 1*(1), 41-64.

Forgas, J. P. (2006). Affective influences on interpersonal behavior: Towards understanding the role of affect in everyday interactions. In J. P. Forgas (Ed.), *Affect in social thinking and behavior* (pp. 269-290). New York: Psychology Press.

Foucault, M. (1978). *The history of sexuality: An introduction* (R. Huxley, Trans.). London: Penguin.

Freeman, L. C. (1979). Centrality in social networks. *Social Networks, 1*, 215-239.

Gardner, W., Mulvey, E. P., & Shaw, E. C. (1995). Regression analyses of counts and rates: Poisson, overdispersed poisson, and negative binomial models. *Psychological Bulletin, 118*, 392-392.

Garfield, E. (1979). *Citation indexing: Its theory and application in science, technology, and humanities*. New York: Wiley.

Garton, L., Haythornthwaite, C., & Wellman, B. (1999). Studying online social networks. *Journal of Computer-Mediated Communication, 3*(1).

Gastwirth, J. L. (1972). The estimation of the Lorenz curve and Gini index. *Review of Economics and Statistics, 54*(3), 306-316.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.

George, J. M. (2000). Emotions and leadership: The role of emotional intelligence. *Human Relations, 53*(8), 1027.

Gibbons, P., Busch, J., & Bradac, J. J. (1991). Powerful versus powerless language: Consequences for persuasion, impression formation, and cognitive response. *Journal of Language and Social Psychology, 10*(2), 115-133.

Gill, A. (2004). *Personality and language: The projection and perception of personality in computer-mediated communication*. University of Edinburgh, Edinburgh.

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

Goffman, E. (1990). *The presentation of self in everyday life*. London: Penguin.

Gould, R. V., & Fernandez, R. (1989). Structures of mediation: A formal approach to brokerage in transaction networks. *Sociological Methodology, 19*(1989), 89-126.

Granovetter, M. (1978). Threshold models of collective behavior. *The American Journal of Sociology, 83*(6), 1420-1443.

Gruhl, D., Guha, R., Liben-Nowell, D., & Tomkins, A. (2004). *Information diffusion through blogspace*. Paper presented at the Proceedings of the 13th international conference on World Wide Web.

Guadagno, R., & Cialdini, R. (2005). Online persuasion and compliance: Social influence on the Internet and beyond. In Y. Amichai-Hamburger (Ed.), *The Social Net: Understanding human behavior in cyberspace* (pp. 91-114). Cambridge: Oxford University Press.

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis* (3rd ed.). New York: Macmillan.

Hamilton, M. A., & Hunter, J. E. (1998). The effect of language intensity on receiver evaluations of message, source, and topic. In M. Allen & R. W. Preiss (Eds.), *Persuasion: Advances Through Meta-analysis* (pp. 99-138). Cresskill, New Jersey: Hampton Press, Inc.

Hampton, K. N., & Wellman, B. (1999). Netville online and offline: Observing and surveying a wired suburb. *American Behavioral Scientist, 43*, 475-492.

Hancock, J. T., Curry, L., Goorha, S., & Woodworth, M. (2005). *Automated linguistic analysis of deceptive and truthful synchronous computer-mediated communication.* Paper presented at the Hawaii International Conference on System Sciences, Hawaii.

Hanneman, R. A., & Riddle, M. (2005). *Introduction to social network methods.* . Riverside, CA: University of California, Riverside.

Hare, A. P. (1976). *Handbook of small group research*. New York: Free Press of Glencoe.

Hart, R. (2000). Diction 5.0: The text analysis program. Austin, Texas: Digitext, Inc.

Hart, R. P. (1984). *Verbal style and the presidency: A computer-based analysis*. New York: Academic Press.

Hart, R. P. (1987). *The sound of leadership: Presidential communication in the modern age*. Chicago: University of Chicago Press.

Hart, R. P., & Childers, J. (2004). Verbal certainty in American politics: An overview and extension. *Presidential Studies Quarterly, 34*(3), 516-535.

Hayes, A. F. (2006). A primer on multilevel modeling. *Human Communication Research, 34*(4), 385-410.

Hemetsberger, A. (2001). *Fostering cooperation on the Internet: Social exchange processes in innovative virtual consumer communities.* Paper presented at the Annual Conference of the Association for Consumer Research (ACR), Austin, Texas.

Herring, S., Job-Sluder, K., Scheckler, R., & Barab, S. (2002). Searching for safety online: Managing" trolling" in a feminist forum. *The Information Society, 18*(5), 371-384.

Herring, S. C. (2002). Computer-Mediated Communication on the Internet. *Annual Review of Information Science and Technology, 36*(1), 109-168.

Herring, S. C. (2003). Gender and power in online communication. In J. Holmes & M. Meyerhoff (Eds.), *The Handbook of Language and Gender* (pp. 202-228). Oxford: Blackwell Publishers.

Hoffmann, S. (2007). Processing internet-derived text -- Creating a corpus of usenet messages. *Literary and Linguistic Computing, 22*(2), 151-165.

Hofmann, D. A., & Gavin, M. B. (1998). Centering decisions in hierarchical linear models: Implications for research in organizations. *Journal of Management, 24*(5), 623-641.

Hollander, E. P. (1961). Emergent leadership and social influence. In L. Petrullo & B. M. Bass (Eds.), *Leadership and interpersonal behavior* (pp. 30-47). New York: Holt, Rinehart and Winston, inc.

Hollander, E. P. (1986). On the central role of leadership processes. *Applied Psychology, 35*(1), 39-52.

Holtgraves, T., & Lasky, B. (1999). Linguistic power and persuasion. *Journal of Language and Social Psychology, 18*(2), 196-205.

Horrigan, J. (2001). *Online communities: Networks that nurture long-distance relationships and local ties*. Washington D.C. : PEW Internet and Family Life Project.

Hosman, L. A. (2002). Language and Persuasion. In J. P. Dillard & M. Pfau (Eds.), *The Persuasion Handbook: developments in theory and practice* (pp. 371-390). New York: Sage.

Hoxx, J. J. (1995). *Applied multilevel analysis*. Amsterdam: TT-Publikaties.

Huberman, B. A., & Adamic, L. A. (2003). Information dynamics in the networked world. *Lecture Notes in Physics, 650*, 371-398.

Huffaker, D., Jorgensen, J., Iacobelli, F., Tepper, P., & Cassell, J. (2006, June 8). *Computational measures for language similarity across time in online communities.* Paper presented at the Workshop on Analyzing Conversations in Text and Speech (ACTS) at Human Language Technologies - North American chapter of the Association for Computational Linguistics (HLT-NAACL), New York, NY.

Hummel, H. G. K., Burgos, D., Tattersall, C., Brouns, F., Kurvers, H., & Koper, R. (2005). Encouraging contributions in learning networks using incentive mechanisms. *Journal of Computer Assisted Learning, 21*(5), 355-365.

Humphrey, R. H. (2002). The many faces of emotional leadership. *The Leadership Quarterly, 13*(5), 493-504.

Jago, A. G. (1982). Leadership: Perspectives in Theory and Research. *Management Science, 28*(3), 315-336.

Jentsch, F., & Smith-Jentsch, K. A. (2001). Assertiveness and team performance: More than "just say no". In E. Salas, C. A. Bowers & E. Edens (Eds.), *Improving Teamwork in Organizations: Applications of Resource Management Training* (pp. 73-94). Mahwah, NJ: Lawrence Erlbaum.

Joyce, E., & Kraut, R. (2006). Predicting continued participation in newsgroups. *Journal of Computer-mediated Communication, 11*(3).

Jurafsky, D., & Martin, J. H. (2000). *Speech and language processing: An introduction to natural language processing, computational linguistics and speech recognition*. New York: Prentice-Hall.

Kakwani, N. C. (1977). Applications of Lorenz curves in economic analysis. *Econometrica, 45*(3), 719-727.

Katz, E., & Lazersfeld, P. F. (1955). *Personal influence: The part played by people in the flow of mass communications*. Glencoe, Illinois: Free Press.

Kelly, J., Fisher, D., & Smith, M. (2005). *Debate, division, and diversity: Political discourse networks in USENET newsgroups*. Paper presented at the Online Deliberation Conference.

Kiesler, S., & Sproull, L. (1992). Group decision making and communication technology. *Organizational Behavior and Human Decision Processes, 52*(1), 96-123.

Kilgarriff, A. (2001). Comparing corpora. *International Journal of Corpus Linguistics, 6*(1), 97-133.

Kittur, A., & Kraut, R. (2008, November 8-12). *Harnessing the wisdom of crowds in Wikipedia: Quality through coordination.* Paper presented at the ACM Conference on Computer-Supported Cooperative Work (CSCW), San Diego, CA.

Kouzes, R. T., Myers, J. D., & Wulf, W. A. (1996). Collaboratories: Doing science on the Internet. *Computer, 29*(8), 40-46.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). Introduction to latent semantic analysis. *Discourse Processes, 25*, 259-284.

Latane, B. (1981). The psychology of social impact. *American Psychologist, 36*(4), 343-356.

Lewis, K. M. (2000). When Leaders Display Emotion: How Followers Respond to Negative Emotional Expression of Male and Female Leaders. *Journal of Organizational Behavior, 21*(2), 221-234.

Ling, K., Beenen, G., Ludford, P., Wang, X., Chang, K., Li, X., et al. (2005). Using social psychology to motivate contributions to online communities. *Journal of Computer-mediated Communication, 10*(4).

Maheswaran, D., & Meyers-Levy, J. (1990). The influence of message framing and issue involvement. *Journal of Marketing Research, 27*(3), 361-367.

Malvern, D., Richards, B., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development: Quantification and assessment*. New York: Palgrave Macmillan.

Manning, C. D., & Shutze, H. (2002). *Foundations of statistical natural language processing*. Cambridge, MA: MIT Press.

Matsumura, N., Ohsawa, Y., & Ishizuka, M. (2002a). *Influence diffusion model in text-based communication*. Paper presented at the International World Wide Web Conference.

Matsumura, N., Ohsawa, Y., & Ishizuka, M. (2002b, September 16-18). *Mining and characterizing opinion leaders from threaded online discussions.* Paper presented at the Knowledge-Based Intelligent Information and Engineering Systems, Crema, Italy.

Matsumura, N., & Sasaki, Y. (2007). Human influence network for understanding leadership behavior. *International Journal of Knowledge-based and Intelligent Engineering Systems, 11*, 291-300.

Matsumura, N., Yamamoto, H., & Tomozawa, D. (2008, March 30 - April 2). *Finding influencers and consumer insights in the blogosphere.* Paper presented at the International Conference on Weblogs and Social Media, Seattle, Washington.

McKee, G., Malvern, D., & Richards, B. (2000). Measuring vocabulary diversity using dedicated software. *Literary and Linguistic Computing, 15*(3), 323-338.

McLeod, P. L., Baron, R. S., Marti, M. W., & Yoon, K. (1997). The eyes have it: Minority influence in face-to-face and computer-mediated group discussion. *Journal of Applied Psychology, 82*(5), 706-718.

McLure Wasko, M., & Faraj, S. (2000). "It is what one does": why people participate and help others in electronic communities of practice. *Journal of strategic information systems, 9*, 155-173.

Mehra, A., Dixon, A. L., Brass, D. J., & Robertson, B. (2006). The social network ties of group leaders: Implications for group performance and leader reputation. *Organization Science, 17*(1), 64.

Menard, S. (1995). Applied logistic regression analysis. Quantitative applications in the social sciences no. 106: Sage Publications, Thousand Oaks, California.

Miller, G. A., Leacock, C., Tengi, R., & Bunker, R. T. (1993). A semantic concordance. *Proceedings of the workshop on Human Language Technology*, 303-308.

Misiolek, N. L., & Heckman, R. (2005, January 3 - 6, 2005). *Patterns of emergent leadership in virtual teams.* Paper presented at the 2005 Hawaii International Conference on System Sciences (HICSS-38), Waikoloa, HI.

Mizruchi, M. S., & Potts, B. B. (1998). Centrality and power revisited: actor success in group decision making. *Social Networks, 20*(4), 353-387.

Monge, P. R., & Contractor, N. (2003). *Theories of communication networks*. New York: Oxford University Press.

Newcombe, M. J., & Ashkanasy, N. M. (2002). The role of affect and affective congruence in perceptions of leaders: an experimental study. *The Leadership Quarterly, 13*(5), 601-614.

Ng, S. H., & Bradac, J. J. (1993). *Power in language: Verbal communication and social influence*. ENWbury Park: Sage Publications.

Niederhoffer, K. G., & Pennebaker, J. W. (2002). Linguistic style matching in social interaction. *Journal of Language and Social Psychology, 21*(4), 337-360.

Nonnecke, B., Andrews, D., & Preece, J. (2006). Non-public and public online community participation: Needs, attitudes and behavior. *Electronic Commerce Research, 6*(1), 7-20.

O'Keefe, D. J. (2002). *Persuasion: Theory and research* (2nd ed.). Thousand Oaks: Sage Publications.

Paolillo, J. (2001). Language variation on internet relay chat: A social network approach. *Journal of Sociolinguistics, 5*(2), 180-213.

Pederson, T., Patwardhan, S., Banerjee, S., & Michelizzi, J. (2008). Text::Similarity. Duluth, Minnesota: University of Minnesota.

Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2006). Linguistic inquiry and word count: LIWC. Austin, Texas: Erlbaum Publishers.

Pennebaker, J. W., & Francis, M. E. (1996). Cognitive, emotional and language processes in disclosure. *Cognition and Emotion, 10*, 601-626.

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology, 54*, 547-577.

Pittinsky, M. S. (Ed.). (2003). *The wired tower: Perspectives on the impact of the internet on higher education*. Upper Saddle River, NJ: Prentice Hall.

Pornpitakpan, C. (2004). The persuasiveness of source credibility: A critical review of five decades' evidence. *Journal of Applied Social Psychology, 34*(2), 243-281.

Porter, M. E. (2001). Strategy and the Internet. *Harvard Business Review, 79*(3), 62-79.

Postmes, T., Spears, R., Sakhel, K., & de Groot, D. (2001). Social Influence in Computer-Mediated Communication: The Effects of Anonymity on Group Behavior. *Personality and Social Psychology Bulletin, 27*(10), 1243-1254.

Preece, J. (2000). *Online communities: Designing usability, supporting sociability*. Chichester, UK: John Wiley and Sons.

Putnam, R. D. (2000). *Bowling alone: The collapse and revival of American community*. New York: Simon & Schuster.

Rainie, L., & Horrigan, J. (2004). *A decade of adoption: How the Internet has woven itself into American life*. Washington D.C.: PEW Internet and Family Life.

Ramachandran, A., & Feamster, N. (2006, September 11-15). *Understanding the network-level behavior of spammers.* Paper presented at the Proceedings of the 2006 Conference on Applications, Technologies, Architectures and Protocols for Computer Communications, Pisa, Italy.

Rashotte, L. (2007). Social Influence. In G. Ritzer (Ed.), *The Blackwell Encyclopedia of Sociology Online* (pp. 4426-4429). Hoboken, NJ: Blackwell.

Raudenbush, S. W. (2004). *HLM 6: Hierarchical Linear and Nonlinear Modeling*. Lincolnwood, IL: Scientific Software International

Ren, Y., Harper, M., Drenner, S., Kiesler, S., Terveen, L., & Riedl, J. (under review). Building attachment to online communities: Design principles for identity and bonds.

Resnick, P., Kuwabara, K., Zeckhauser, R., & Friedman, E. (2000). Reputation Systems. *Communications of the ACM, 43,* 45-48.

Rheingold, H. (2000). *The virtual community: Homesteading on the electronic frontier*. Cambridge: M.I.T. Press.

Rheingold, H. (2003). *Smart mobs: The next social revolution*. Cambridge: Perseus Publishing.

Riggio, R. E., Riggio, H. R., Salinas, C., & Cole, E. J. (2003). The role of social and emotional communication skills in leader emergence and effectiveness. *Group Dynamics: Theory Research and Practice, 7*(2), 83-103.

Rogers, E. M. (1995). *The diffusion of innovation* (Fourth ed.). New York: Free Press.

Rosen, K. T., & Resnick, M. (1979). The size distribution of cities: An examination of the Pareto law and primacy. *Journal of Urban Economics, 8*(2), 165-186.

Ryan, S., Scott, B., Freeman, H., & Patel, D. (2000). *The virtual university: The internet and resource-based learning*. New York, NY: Routledge.

Salovey, P. (1990). Emotional intelligence. *Imagination, Cognition and Personality, 9*(3).

Sapir, E. (1949). *Culture, language and personality*. Berkeley: University of California Press.

Sarker, S., & Grewel, R. (2002). *Emergence of leaders in virtual teams: What matters?* Paper presented at the 35th Annual Hawaii International Conference on System Sciences (HCISS'02), Honolulu, Hawaii.

Schegloff, E. A. (1998). Talk and social structure. In J. Coates (Ed.), *Language and Gender: A Reader* (pp. 107-120). Malden, MA: Blackwell Publishers Ltd.

Slater, M. D., Snyder, L., & Hayes, A. F. (2006). Thinking and modeling at multiple levels: The potential contribution of multilevel modeling to communication theory and research. *Human Communication Research, 32*(4), 375-384.

Smith, M. A. (1999). Invisible crowds in cyberspace: Mapping the social structure of usenet. In M. Smith & P. Kollock (Eds.), *Communities in Cyberspace, edited by* (pp. 194-218). London: Routledge Press.

Smith, M. A., Researcher at Microsoft Corporation (2007). Netscan: A social accounting search engine: Microsoft Research Community Technologies.

Smith, S. M., & Petty, R. E. (1996). Message framing and persuasion: A message processing analysis. *Personality and Social Psychology Bulletin, 22*(3), 257-268.

Snijders, T. A. B., Pattison, P. E., Robins, G. L., & Handcock, M. S. (2006). New specifications for exponential random graph models. *Sociological Methodology, 36*(1), 99-153.

Sorrentino, R. M., & Boutillier, R. G. (1975). The effect of quantity and quality of verbal interaction on ratings of leadership ability. *Journal of Experimental Social Psychology, 11*(5), 403-411.

Sparrowe, R. T., Liden, R. C., Wayne, S. J., & Kraimer, M. L. (2001). Social networks and the performance of individuals and groups. *Academy of Management Journal, 44*(2), 316-325.

Sproull, L., & Kiesler, S. (1991). *Connections: New ways of working in the networked organization*. Cambridge: MIT Press.

Sternberg, R. J. (1999). Successful intelligence: finding a balance. *Trends in Cognitive Sciences, 3*(11), 436-442.

Sudweeks, F., & Simoff, S. (2005, January 3-6). *Leading conversations: Communication behaviours of emergent leaders in virtual teams.* Paper presented at the Thirty-Eighth Annual Hawaii International Conference on System Sciences (HICSS-38), Big Island, Hawaii.

Suler, J. (2004). The online disinhibition effect. *CyberPsychology & Behavior, 7*(3), 321-326.

Tepper, M. (1997). Usenet communities and the cultural politics of information. In D. Porter (Ed.), *Internet Culture* (pp. 39-54). New York: Routledge.

Thayer, L. (1988). Leadership/Communication: A critical review and a modest proposal. In G. M. Goldhaber & G. A. Barnett (Eds.), *Handbook of organizational communication* (pp. 231-263). Norwood, New Jersey: Ablex Publishing Corporation.

Turner, F. (2005). Where the counterculture met the new economy: The WELL and the origins of virtual community. *Technology and Culture, 46*(3), 485-512.

Turner, J. C. (1991). *Social influence*. Buckingham: Open University Press.

Turner, T. C., Smith, M. A., Fisher, D., & Welser, H. T. (2005). Picturing usenet: Mapping computer-mediated collective action. *Journal of Computer-Mediated Communication, 10*(4).

Valente, T. W., & Davis, R. L. (1999). Accelerating the diffusion of innovations using opinion leaders. *The Annals of the American Academy Of Political And Social Science, 566*, 55-67.

Wallace, P. (2004). *The internet in the workplace: How new technology is transforming work.* Cambridge: Cambridge University Press.

Walther, J. B. (1996). Computer-mediated communication: Impersonal, interpersonal and hyperpersonal Interaction. *Communication Research, 23*(1), 3-43.

Wang, X., Kraut, R., Butler, B., Burke, M., & Joyce, E. (under review). Beyond information: Developing the relationship between the individual and the group in online communities. *Information Systems Research*.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge: Cambridge University Press.

Watts, D. J. (2002). A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences, 99*(9), 5766.

Watts, D. J., & Dodds, P. S. (2007). Influentials, networks, and public opinion formation. *Journal of Consumer Research, 34*(4), 441-458.

Weber, M. (1947). *The theory of social and economic organization* (A. M. Henderson & T. Parsons, Trans.). New York: Oxford University Press.

Weimann, G. (1994). *The influentials: People who influence people.* Albany: State University of New York Press.

Weisband, S. (1994). Overcoming social awareness in computer-supported groups. *Computer Supported Collaborative Work (CSCW), 2*(4), 1573-7551.

Weisband, S. P., Schneider, S. K., & Connolly, T. (1995). Computer-mediated communication and social information: Status salience and status differences. *The Academy of Management Journal, 38*(4), 1124-1151.

Wellman, B., & Gulia, M. (1999). Net surfers don't ride alone: Virtual communities as communities. In P. Kollock & M. Smith (Eds.), *Communities in Cyberspace*. New York: Routledge.

Wellman, B., Salaff, J., Dimitrova, D., Garton, L., Gulia, M., & Haythornthwaite, C. (1996). Computer networks as social networks: Collaborative work, telework, and virtual community *Annual Review of Sociology*, 213-238.

Wittgenstein, L. (1958). *Philosophical investigations* (G. E. M. Anscombe, Trans. 3 ed.). Englewood Cliffs, NJ: Prentice-Hall.

Yoo, Y., & Alavi, M. (2004). Emergent leadership in virtual teams: what do emergent leaders do? *Information and Organization, 14*(1), 27-58.

Zaccaro, S. J., Kemp, C., & Bader, P. (2004). Leader traits and attributes. In J. Antonakis, A. T. Cianciolo & R. J. Sternberg (Eds.), *The nature of leadership*. Thousand Oaks, CA: Sage Publications.

Appendix A

Stop Words Removed for Language Diffusion Analysis

a	become	eight	hasnt	many	only
about	becomes	either	have	may	onto
above	becoming	eleven	he	me	or
across	been	else	hence	meanwhile	other
after	before	elsewhere	her	might	others
afterwards	beforehand	empty	here	mill	otherwise
again	behind	enough	hereafter	mine	our
against	being	etc	hereby	more	ours
all	below	even	herein	moreover	ourselves
almost	beside	ever	hereupon	most	out
alone	besides	every	hers	mostly	over
along	between	everyone	herself	move	own
already	beyond	everything	him	much	part
also	bill	everywhere	himself	must	per
although	both	except	his	my	perhaps
always	bottom	few	how	myself	please
am	but	fifteen	however	name	put
among	by	fify	hundred	namely	rather
amongst	call	fill	i	neither	re
amoungst	can	find	ie	never	same
amount	cannot	fire	if	nevertheless	see
an	cant	first	in	next	seem
and	co	five	inc	nine	seemed
another	computer	for	indeed	no	seeming
any	con	former	interest	nobody	seems
anyhow	could	formerly	into	none	serious
anyone	couldnt	forty	is	noone	several
anything	cry	found	it	nor	she
anyway	de	four	its	not	should
anywhere	describe	from	itself	nothing	show
are	detail	front	keep	now	side
around	do	full	last	nowhere	since
as	done	further	latter	of	sincere
at	down	get	latterly	off	six
back	due	give	least	often	sixty
be	during	go	less	on	so
became	each	had	ltd	once	some
because	eg	has	made	one	somehow

someone
something
sometime
sometimes
somewhere
still
such
system
take
ten
than
that
the
their
them
themselves
then
thence
there
thereafter
thereby
therefore
therein
thereupon
these
they
thick
thin
third
this
those
though
three
through
throughout
thru
thus
to
together
too
top
toward
towards
twelve
twenty
two
un
under
until
up
upon
us
very
via
was
we
well
were
what
whatever
when
whence
whenever
where
whereafter
whereas
whereby
wherein
whereupon
wherever
whether
which
while
whither
who
whoever
whole
whom
whose
why
will
with
within
without
would
yet
you
your
yours
yourself
yourselves

Appendix B

Words Used in Linguistic Inquiry and Word Count (LIWC) Dictionaries

Affect

abandon*	anguish*	benign*	clever*	cut	disadvantage*
abuse*	annoy*	best	comed*	cute*	disagree*
abusi*	antagoni*	better	comfort*	cutie*	disappoint*
accept	anxi*	bitch*	commitment*	cynic	disaster*
accepta*	aok	bitter*	compassion*	damag*	discomfort*
accepted	apath*	blam*	complain*	damn*	discourag*
accepting	appall*	bless*	compliment*	danger*	disgust*
accepts	appreciat*	bold*	concerned	daring	dishearten*
ache*	apprehens*	bonus*	confidence	darlin*	disillusion*
aching	argh*	bore*	confident	daze*	dislike
active*	argu*	boring	confidently	dear*	disliked
admir*	arrogan*	bother*	confront*	decay*	dislikes
ador*	asham*	brave*	confus*	defeat*	disliking
advantag*	assault*	bright*	considerate	defect*	dismay*
adventur*	asshole*	brillian*	contempt*	defenc*	dissatisf*
advers*	assur*	broke	contented*	defens*	distract*
affection*	attachment*	brutal*	contentment	definite	distraught
afraid	attack*	burden*	contradic*	definitely	distress*
aggravat*	attract*	calm*	convinc*	degrad*	distrust*
aggress*	aversi*	care	cool	delectabl*	disturb*
agitat*	avoid*	cared	courag*	delicate*	divin*
agoniz*	award*	carefree	crap	delicious*	domina*
agony	awesome	careful*	crappy	deligh*	doom*
agree	awful	careless*	craz*	depress*	dork*
agreeab*	awkward*	cares	create*	depriv*	doubt*
agreed	bad	caring	creati*	despair*	dread*
agreeing	bashful*	casual	credit*	desperat*	dull*
agreement*	bastard*	casually	cried	despis*	dumb*
agrees	battl*	certain*	cries	destroy*	dump*
alarm*	beaten	challeng*	critical	destruct*	dwell*
alone	beaut*	champ*	critici*	determina*	dynam*
alright*	beloved	charit*	crude*	determined	eager*
amaz*	benefic*	charm*	cruel*	devastat*	ease*
amor*	benefit	cheat*	crushed	devil*	easie*
amus*	benefits	cheer*	cry	devot*	easily
anger*	benefitt*	cherish*	crying	difficult*	easiness
angr*	benevolen*	chuckl*	cunt*	digni*	easing
easy*	fears	fun	gross*	homesick*	insult*

ecsta*	feroc*	funn*	grouch*	honest*	intell*
efficien*	festiv*	furious*	grr*	honor*	interest*
egotis*	feud*	fury	guilt*	honour*	interrup*
elegan*	fiery	geek*	ha	hope	intimidat*
embarrass*	fiesta*	genero*	haha*	hoped	invigor*
emotion	fight*	gentle	handsom*	hopeful	irrational*
emotion	fine	gentler	happi*	hopefully	irrita*
emotional	fired	gentlest	happy	hopefulness	isolat*
empt*	flatter*	gently	harass*	hopeless*	jaded
encourag*	flawless*	giggl*	harm	hopes	jealous*
enemie*	flexib*	giver*	harmed	hoping	jerk
enemy*	flirt*	giving	harmful*	horr*	jerked
energ*	flunk*	glad	harming	hostil*	jerks
engag*	foe*	gladly	harmless*	hug	joke*
enjoy*	fond	glamor*	harmon*	hugg*	joking
enrag*	fondly	glamour*	harms	hugs	joll*
entertain*	fondness	gloom*	hate	humiliat*	joy*
enthus*	fool*	glori*	hated	humor*	keen*
envie*	forbid*	glory	hateful*	humour*	kidding
envious	forgave	goddam*	hater*	hurra*	kill*
envy*	forgiv*	good	hates	hurt*	kind
evil*	fought	goodness	hating	ideal*	kindly
excel*	frantic*	gorgeous*	hatred	idiot	kindn*
excit*	freak*	gossip*	hazy	ignor*	kiss*
excruciat*	free	grace	heartbreak*	immoral*	laidback
exhaust*	freeb*	graced	heartbroke*	impatien*	lame*
fab	freed*	graceful*	heartfelt	impersonal	laugh*
fabulous*	freeing	graces	heartless*	impolite*	lazie*
fail*	freely	graci*	heartwarm*	importan*	lazy
faith*	freeness	grand	heaven*	impress*	liabilit*
fake	freer	grande*	heh*	improve*	liar*
fantastic*	frees*	gratef*	hell	improving	libert*
fatal*	friend*	grati*	hellish	inadequa*	lied
fatigu*	fright*	grave*	helper*	incentive*	lies
fault*	frustrat*	great	helpful*	indecis*	like
favor*	fuck	greed*	helping	ineffect*	likeab*
favour*	fucked*	grief	helpless*	inferior*	liked
fear	fucker*	griev*	helps	inhib*	likes
feared	fuckin*	grim*	hero*	innocen*	liking
fearful*	fucks	grin	hesita*	insecur*	livel*
fearing	fume*	grinn*	hilarious	insincer*	LMAO
fearless*	fuming	grins	hoho*	inspir*	LOL
lone*	missing	outgoing	prais*	respect	shared
longing*	mistak*	outrag*	precious*	restless*	shares
lose	mock	overwhelm*	prejudic*	revenge*	sharing
loser*	mocked	pain	pressur*	revigor*	shit*
loses	mocker*	pained	prettie*	reward*	shock*

losing
loss*
lost
lous*
love
loved
lovely
lover*
loves
loving*
low*
loyal*
luck
lucked
lucki*
luckless*
lucks
lucky
ludicrous*
lying
mad
maddening
madder
maddest
madly
magnific*
maniac*
masochis*
melanchol*
merit*
merr*
mess
messy
miser*
miss
missed
misses
stink*
strain*
strange
strength*
stress*
strong*
struggl*
stubborn*
stunk
stunned

mocking
mocks
molest*
mooch*
mood
moodi*
moods
moody
moron*
mourn*
murder*
nag*
nast*
neat*
needy
neglect*
nerd*
nervous*
neurotic*
nice*
numb*
nurtur*
obnoxious*
obsess*
offence*
offend*
offens*
ok
okay
okays
oks
openminded*
openness
opportun*
optimal*
optimi*
original
sweet
sweetheart*
sweetie*
sweetly
sweetness*
sweets
talent*
tantrum*
tears
teas*

painf*
paining
painl*
pains
palatabl*
panic*
paradise
paranoi*
partie*
party*
passion*
pathetic*
peace*
peculiar*
perfect*
personal
perver*
pessimis*
petrif*
pettie*
petty*
phobi*
piss*
piti*
pity*
play
played
playful*
playing
plays
pleasant*
please*
pleasing
pleasur*
poison*
popular*
positiv*
treasur*
treat
trembl*
trick*
trite
triumph*
trivi*
troubl*
true
trueness

pretty
prick*
pride
privileg*
prize*
problem*
profit*
promis*
protest
protested
protesting
proud*
puk*
punish*
radian*
rage*
raging
rancid*
rape*
raping
rapist*
readiness
ready
reassur*
rebel*
reek*
regret*
reject*
relax*
relief
reliev*
reluctan*
remorse*
repress*
resent*
resign*
resolv*
valuabl*
value
valued
values
valuing
vanity
vicious*
victim*
vigor*
vigour*

rich*
ridicul*
rigid*
risk*
ROFL
romanc*
romantic*
rotten
rude*
ruin*
sad
sadde*
sadly
sadness
safe*
sarcas*
satisf*
savage*
save
scare*
scaring
scary
sceptic*
scream*
screw*
secur*
selfish*
sentimental*
serious
seriously
seriousness
severe*
shake*
shaki*
shaky
shame*
share
wisdom
wise*
witch
woe*
won
wonderf*
worr*
worse*
worship*
worst

shook
shy*
sicken*
sigh
sighed
sighing
sighs
silli*
silly
sin
sincer*
sinister
sins
skeptic*
slut*
smart*
smil*
smother*
smug*
snob*
sob
sobbed
sobbing
sobs
sociab*
solemn*
sorrow*
sorry
soulmate*
special
spite*
splend*
stammer*
stank
startl*
steal*
stench

stuns	tehe	truer	vile	worthless*
stupid*	temper	truest	villain*	worthwhile
stutter*	tempers	truly	violat*	wow*
submissive*	tender*	trust*	violent*	wrong*
succeed*	tense*	truth*	virtue*	yay
success*	tensing	turmoil	virtuo*	yays
suck	tension*	ugh	vital*	yearn*
sucked	terribl*	ugl*	vulnerab*	
sucker*	terrific*	unattractive	vulture*	
sucks	terrified	uncertain*	war	
sucky	terrifies	uncomfortabl*	warfare*	
suffer	terrify	uncontrol*	warm*	
suffered	terrifying	uneas*	warred	
sufferer*	terror*	unfortunate*	warring	
suffering	thank	unfriendly	wars	
suffers	thanked	ungrateful*	weak*	
sunnier	thankf*	unhapp*	wealth*	
sunniest	thanks	unimportant	weapon*	
sunny	thief	unimpress*	weep*	
sunshin*	thieve*	unkind	weird*	
super	thoughtful*	unlov*	welcom*	
superior*	threat*	unpleasant	well*	
support	thrill*	unprotected	wept	
supported	ticked	unsavo*	whine*	
supporter*	timid*	unsuccessful*	whining	
supporting	toleran*	unsure*	whore*	
supportive*	tortur*	unwelcom*	wicked*	
supports	tough*	upset*	willing	
suprem*	traged*	uptight*	wimp*	
sure*	tragic*	useful*	win	
surpris*	tranquil*	useless*	winn*	
suspicio*	trauma*	vain	wins	

*Note: 915 words are included in this dictionary. *includes any derivation of the root word.*

Certainty

absolute	correct*	fundamentalis*	prove*
absolutely	defined	fundamentally	pure*
accura*	definite	fundamentals	sure*
all	definitely	guarant*	total
altogether	definitive*	implicit*	totally
always	directly	indeed	true
apparent	distinct*	inevitab*	truest
assur*	entire*	infallib*	truly
blatant*	essential	invariab*	truth*
certain*	ever	irrefu*	unambigu*

clear	every	must	undeniab*
clearly	everybod*	mustnt	undoubt*
commit	everything*	must'nt	unquestion*
commitment*	evident*	mustn't	wholly
commits	exact*	mustve	
committ*	explicit*	must've	
complete	extremely	necessar*	
completed	fact	never	
completely	facts	obvious*	
completes	factual*	perfect*	
confidence	forever	positiv*	
confident	frankly	precis*	
confidently	fundamental	proof	

*Note: 83 words are included in this dictionary. *includes any derivation of the root word.*