

Beyond Information: Developing the Relationship between the Individual and the Group in Online Communities

Robert Kraut¹, Xiaoqing Wang², Brian Butler³, Elisabeth Joyce⁴ and Moira Burke¹

¹Carnegie Mellon University
Pittsburgh, PA 15213

²University of Maryland
College Park, MD 20742

³University of Pittsburgh
Pittsburgh, PA 15260

⁴Edinboro University
Edinboro, PA 16444

April 18, 2008

Acknowledgments:

We thank Zoe Ouyang, Kenneth Chan, and David J. Housman for research assistance; Marc A. Smith at Microsoft Research and David Jeske at Google for providing data; Carolyn Rosé, Jaime Arguello, Cameron Williams and William Cohen at Carnegie Mellon University for advice about text and content analysis; and Kimberly Ling for helpful feedback and discussion throughout the process of conducting this search. This research is supported by NSF grant IIS-0325049.

Beyond Information: Developing the Relationship between the Individual and the Group in Online Communities

Abstract

Online communities are increasingly important for both the businesses and the general public that uses them. However, current IS research on online communities has a limited view of these groups, treating them primarily as information repositories, where people participate by seeking or contributing knowledge. This paper argues that online communities are also social systems with which people form relationships and repeatedly interact over time. This paper proposes a relational model of online communities that emphasizes the social, bi-directional, and dynamic nature of the interactions in online communities and the way that these interactions build and maintain the relationship between individuals and the community. Hypotheses derived from the relational model are tested by examining the 22-month history of 28,869 newcomers who initially posted to 98 Usenet groups between May 15, 2003, and February 23, 2005. First, taking the newcomers' point of view, we examine how the community's response to their posts influences their subsequent participation in the community. Then, taking the community's point of view, we examine how the nature of individual members' participation in the community influences the community's willingness to interact with them. Consistent with the relational model, the results show that the social nature of interactions, including the social status of the repliers and the use of welcoming, inclusive language in community responses, facilitates the development of the relationship between newcomers and the community. Similarly, newcomers' prior experience in replying to messages and their use of self-introductions in initial messages, signaling connection to the group, increase the community's willingness to maintain a relationship with them. The findings hold implications for online community researchers and practitioners interested in fostering member engagement. We suggest applying a relational model to studying not only online communities, but also information systems that are increasingly social in nature.

Keywords: Online communities, Virtual communities, Participation, Commitment

Beyond Information: Developing the Relationship between the Individual and the Group in Online Communities

Online communities are large, persistent collections of individuals with common or complementary interests whose primary method of communication is over the Internet. These communities are increasingly important for both businesses and the general public. They are channels through which companies get input from customers, employees and business partners (Wagner 2007), sources of professional and technical information (Lakhani and Hippel 2003, Lakhani and Von Hippel 2003, Wasko and Faraj 2005), platforms for new business models (Verona, et al. 2006), and valuable financial assets (Liedtke 2007). For the general public, online communities provide repositories of useful information (e.g., Wikipedia.org), sources of emotional support for those with illnesses (Davison, et al. 2000), sites for discussion of political and social issues (Hill and Hughes 1998), ways to meet new people and maintain social networks (Wellman 2001), and opportunities for entertainment (Williams, et al. 2006). The growing importance of online communities in many spheres has stimulated the interests of information systems (IS) researchers and professionals in the dynamics and development of technology-enabled communities.

The current IS literature on online communities tends to take an information-centric perspective, treating online communities as places where people seek and exchange information or advice. For example, Wasko and Faraj (2005) describe them as “computer-mediated discussion forums focused on problems of practice that enable individuals to exchange advice and ideas with others based on common interests.” This information-centric view of online communities attempts to explain why participants donate valuable intellectual property to a community without explicit compensation or an employment contract (Constant, et al. 1996, Kankanhalli, et al. 2005, Ma and Agarwal 2007, Roberts, et al. 2006, Wasko and Faraj 2005), the quality of the information produced (Gu, et al. 2007), the impact of the information exchanged (Dellarocas 2006) and problems of information overload (e.g., Butler 2001, Jones, et al. 2004). Even research examining social processes in online communities concentrates on the seeking and provision of information (Ahuja and Galvin 2003).

While the information content of online communities and other instrumental benefits that participants receive from being members are undoubtedly important and provide major motivations for participation, online communities are more than information sources. They are also social systems that share important characteristics of small groups, voluntary associations and organizations (Butler 2004). As is the case with many other groups and organizations, successfully becoming a member of an online community often involves building a relationship with the group as a whole or with the people in it and, over time, becoming committed to that relationship (Sassenberg 2002).

This paper extends previous IS research on online communities by proposing and testing a relational model of online communities, in which online communities are conceived of as social systems whose existence depends upon the relationships between individuals and the community that develop and are maintained through social interactions. The proposed model highlights four elements that have received insufficient attention in the prior IS research on online communities. First is an explicit recognition that the relationship between the individual and group is inherently social and that social factors influence its development over and above informational benefits that people receive from participating in the community. Second, the relationship between a group and an individual is bidirectional. That is, for a newcomer to develop a relationship with an online community, the newcomer must approach the group and participate, and at the same time other members of the community must accept the newcomer (Moreland and Levine 2001). Third, social relationships evolve and change, with different factors influencing them early and later in their histories (Bauer, et al. 2007, Moreland and Levine 2001). Finally, social interaction, not just resource exchange, is significant to the development of individual-community relationships, with the relationship both being manifested and shaped by the interactions (Duck 1988, Eisenberg, et al. 1983).

To develop a relational model, this paper focuses on newcomers to online communities and tests hypotheses about the ways in which social interaction influences the strength of the relationship that they develop with the community. In doing so, we attempt to explain not only why people contribute to online

communities, the focus of much IS research, but also how bi-directional social interactions lead newcomers to become more committed to the community and engage it as an active partner over the long term.

A Relational Model of Online Communities

The relational model of online communities posits that while resource contribution and consumption are important elements of behavior in an online community, they are important especially because of their role in fostering social relationships between the individual and the group (Homans 1958). Social relationships are the connection between two social entities. Studies of interpersonal relationships often focus on relationships between individuals (Berscheid, et al. 1989), while studies of group and organization commitment focus on relationships between an individual and a group or an organization (Mowday 1979). Relationships have both affective and behavioral components. For example, people may feel psychologically connected to a group and act in ways consistent with the group's norms because of this connection (Sassenberg, 2002).

Although relationship is an abstract construct, its strength is generally manifested in observable behavior, such as the frequency, nature, and quality of interaction and its duration (Granovetter 1973). For example, individuals who report greater commitment to a group or organization stay with it longer and contribute more (Bauer, et al. 2007), especially the type of voluntary pro-social contributions known as organizational citizenship behavior (Organ and Ryan 1995). Thus, individuals' relationship with an online community encompasses more than just motivations to contribute resources: it can be reflected in the length of time they stay engaged, the amount they contribute and the types of contributions they make. Similarly, a community's relationship with an individual is likely to be manifested in the willingness of the community to support him or her. In an online community, one manifestation of this support is the willingness of the community to respond to the individual's requests. Within the relational model of online communities, information exchanges and other exchanges of resources are indications that a relationship exists between the individual and the community.

A social relationship between an individual and an online community may be minimal or well-developed and long-lived. Although some visitors to an online community may develop longer-term rela-

tionships with the community, many do not, and the relationship between newcomers and the group is especially fragile. For example, 68% of newcomers to Usenet groups were never seen after their first post (Arguello, et al. 2006), over half of the developers who registered to participate in the Perl open-source development project never returned after posting a single message (Ducheneaut 2005), and about a quarter of the members of guilds in the massive multiplayer game, World of Warcraft, leave their guilds every month even though they are still playing the game (Ducheneaut, et al. 2007). Given these low base rates, understanding the factors that lead to successful relationships between online communities and prospective members will have important theoretical and practical implications.

To understand how relationships between online communities and prospective members evolve, we propose a relational model of online communities based on theories of socialization in small, off-line groups (Moreland and Levine 1988, Moreland and Levine 2001) and commitment in organizations (Bauer, et al. 2007, Jones 1986, Van Maanen and Schein 1979). The relational model inspired by these theories highlights four elements insufficiently addressed in more information-centric views: the roles of social factors, the bi-directional nature of the relationship, development of relationships over time and the role of social interaction.

First is an explicit recognition that the relationship between the individual and community is inherently social, and that social factors influence its development over and above the instrumental, informational benefits people receive from participating in them. Although recent research has begun to examine the influence of social factors in online communities, most has done so while still conceptualizing online communities as information repositories rather than as social entities in their own right (with some exceptions, including Bateman et al, 2006; Blanchard & Markus, 2004). For example, scholars have shown that contributions to online communities are driven by individuals' perceived identity within the community (Ma and Agarwal 2007), organizational climate (Bock, et al. 2005), and the norm of trust and sharing within the community (Bock, et al. 2005, Kankanhalli, et al. 2005). This research examines a variety of social factors, but frames them as contextual elements that influence contribution. As a result, this research rarely looks at the interplay between an individual and a community and particularly, how the

communities interact with the individuals. The relational model of online communities underscores that participation is not only information exchange, but also individuals' attempts to engage a community, which includes effort to send and interpret signals of intention and expectation. For example, patients are more likely to grow committed to an online support group when the group acts in a welcoming manner, independent of the medical information or practical advice it provides (Arguello et al. 2006). Open source development groups may be more willing to accept applicants who espouse the egalitarian values of the open-source movement, independent of the applicants' skills as developers (Stewart and Gosain 2006). Although many social factors can influence the strength of the relationship between individual and group, individuals' attempts to claim membership in an organization and their perceptions that the organization has granted them can be especially important for marginal members, such as remote workers in a distributed organization (Bartel and Dutton 2001, Bartel, et al. 2007). Membership claims are probes through which marginal members assert their membership in a community and assess the acceptance that it provides them. Exchanges of membership claims and membership grants have been shown to lead newcomers to feel higher identification with and commitment to a community (Bartel, et al. 2007).

Second, the relationship between individuals and a community is bi-directional. Models of system use in IS typically examine the factors that cause individuals to accept a system (Venkatesh, et al. 2003) without examining what causes the system to accept the user. Consistent with this tradition, most studies of online communities describe factors that motivate people to contribute to it, but little on how communities accept potential participants. While the uni-directional assumption is defensible when dealing with traditional IS tools (Orlikowski and Iacono 2001), a bi-directional approach is necessary when the 'system' is itself a social entity such as an online community. Unlike the cases typically described by the technology acceptance models, a group must also accept an individual if there is to be continued engagement (Moreland and Levine 2001). An individual-community relationship may fail because a community is insufficiently useful or welcoming, causing the individual to abandon the relationship, or it may fail because the individual demonstrates undesirable behaviors or intentions, causing the community to reject

him or her. Therefore, explaining engagement with online communities requires consideration of both sides of the relationship.

Third, relationships evolve over time. Although substantial research on online communities attempts to explain why individuals differ in their level of participation (Koh, et al. 2007, Wasko and Faraj 2005), little of it examines how the interaction and relationship between individuals and communities change over time. Yet evidence from both off-line organizations (Bauer, et al. 2007) and online groups (Arguello et al, 2006) suggests that exploration processes are especially important early in newcomers' interactions with a group, when they investigate the group and the group evaluates them. For example, in conventional organizations, newcomers' information-seeking influences their subsequent organizational commitment, as they attempt to understand whether existing members of the group will accept them, what benefits they may receive from membership and how they should behave (Bauer, et al. 2007, Miller and Jablin 1991). In addition, organizational attempts to assess newcomers also lead to better person-organizational fit and subsequent commitment (Kristof 1996). The relationship model suggests that both individuals' and group's behavior will evolve over time as the relationship between them develops, and that signals that help each evaluate the other will be especially important early in the relationship.

Fourth, the relational model highlights the social interactions through which individual-community relationships develop. Social interactions are the joint activity or communication in which the parties recognize the presence of others, signal intentions and expectations, and change behavior in response to them. Social interaction is at the core of social relationships (Duck 1998). In online communities, social interactions are both the conduits through which people seek and exchange information or social support (Ridings and Gefen 2004, Wasko and Faraj 2005) and the mechanisms through which individuals and communities learn about the intentions, interests, and expectations of one another (Arguello, et al. 2006). In addition, social interaction is a signal for both the individual and the group that the relationship exists. Whether it is coordinated raiding in the online game World of Warcraft, joint editing in the online encyclopedia Wikipedia, arguments in online political discussion groups, or question-answer exchanges in technical support communities, social interaction is a major component in online communi-

ties (Rafaeli 1988). Therefore, social interaction is essential for the formation and maintenance of a relationship between an individual and an online community.

Hypotheses

The relational model of online communities argues that the information exchange and other behavior observed in online communities is a reflection of an evolving, bi-directional, interaction-based social relationship between an individual and a community. In the sections below, we present hypotheses derived from this model, linking individual and group behaviors with the individual-community relationship.

Relationship development from a newcomer's perspective

In all online communities public and private communication is a major component of the interaction (Rafaeli 1988), and because of the computer-mediated nature of online interaction, text-based conversational exchanges are most common. When individuals make conversational overtures and other community members respond, social interaction has occurred. When people receive responses, they may get useful information or diversion or may find others with similar interests. Regardless of the content of these interactions, however, their mere existence signals that a social relationship exists (Duck, et al. 1991) and is likely to affect an individual's evaluation of the community. For example, Williams and his colleagues demonstrated experimentally that being ignored in even minimal online groups is an aversive experience and causes people to leave the group, while having interaction is perceived as a positive experience and a signal of welcoming, which caused them to stick around (Williams, et al. 2000). Even in cases where individuals seek a debate, argument, or an exchange of insults (Golder and Donath 2004), a community's willingness to respond is central to continuation of the relationship. Therefore, we expect that

H1: The number of responses an individual receives in a social interaction in an online community increases the strength of his or her relationship with the community.

While social interactions should generally affect formation and maintenance of relationships with an online community, individuals will be especially sensitive to social interactions during their early ex-

periences with the community. Group socialization theory, which focuses on groups with clear role transitions between members and non-members (Moreland and Levine 1988, Moreland and Levine 2001), suggests that evaluation and engagement processes are likely to be qualitatively different after people transition from visitors to members (e.g., between a fraternity recruit and a pledge). The organizational socialization literature makes a similar, though tacit, distinction among organizational members, examining how socialization tactics influence organizational commitment among new members, typically recently hired employees.

Even though many conversational communities, including those in this study, do not have a crisp distinction between members and non-members, similar principles apply. Newcomers actively seek information about their fit with a community through their interactions with it, and their knowledge accumulates through participation. Therefore, social interactions embedded in community responses are more informative for newcomers who just started interacting with a community than they would be for those with more experience with the group. Because individuals gain direct experience and knowledge about a group by actively engaging the group, their level of prior participation serves as an indication of how much they know about the group and the extent to which a relationship with the group already exists. Therefore, the impact of community responses is expected to be higher for newcomers when they initially interact with a community than after they have more experience with it.

H2: Number of responses in a social interaction will have a larger impact on individuals' relationship to the community when they first start to participate in it and will decrease as they participate more.

We have argued that newcomers will be more likely to remain engaged with a community if existing members interact with them. However, this reasoning does not differentiate people's experience with an online community *qua* social entity from their experience with it as an information source. If online communities go beyond information exchanges, one would expect that the social character of interactions, signaling the groups' intentions toward newcomers and evaluations of them, will have significant effects on individuals' engagement with the community above and beyond the mere existence of in-

teractions. Here we focus on two social aspects of an interaction: the social status of the repliers and the welcoming character of the repliers' language.

Interacting with a high status member of a community can be a positive experience for an individual. Research on both offline and online groups has shown that individuals with high status tend to have higher influence than those with low status (Berger, et al. 1977, Weisband, et al. 1995) and are more likely to be perceived as credible and competent (Collins and Stukas 2006). As a result, communication with high-status partners tends to be more effective (Hovland and Weiss 1952). This suggests that interacting with members of high status in a community will have a stronger influence on newcomers' relationship development with the community than interacting with low-status members. In addition, the endorsement of the relationship, which is implied by getting a response from the community, is likely to be perceived as a more authoritative reflection of the group's sentiment if it comes from a high-status community member. Therefore, individuals should have higher motivation to develop a strong relationship with a community when they receive responses from high-status members.

Some online communities, like Slashdot.org, maintain explicit reputation systems and leader boards to provide up-to-date views of members' status and levels of contribution. In other communities people must use behavioral cues to assess status. Even though computer-mediated settings hide many social cues (Sproull and Kiesler 1991), people are still able to assess status in these settings (Weisband, et al. 1995). For example, status within a community can be revealed by the behavioral pattern of individuals' electronic communication (Tyler, et al. 2005). Highly active participants who have posted frequently are more visible in the community and are more likely to be seen as high-status than those who post infrequently or have been inactive (Saunders, et al. 1994, Weisband, et al. 1995). As a result, we expect that receiving responses from more active members will lead to a stronger relationship with a community.

H3: Social interaction will have a larger impact on individuals' relationship to the community when responses to their posts come from more active members in the community.

Besides the social status of the responder, the social tone of a response is also likely to influence recipients' relationship with a community. Social interaction in online communities is heavily, if not en-

tirely, embodied in text-based conversation. Unlike offline interactions where people can use a variety of cues to show welcoming or affect, online interaction is mediated by communication technologies and allows limited cues. Therefore, the language used in a message provides the primary source of social signals such as attachment, legitimacy, and friendliness (Galegher, et al. 1998, Lyytinen 1985). When current members of the community use language that signals they already have some connection to an individual, they imply that the recipient is welcomed and that the relationship is desirable. This language represents a form of “membership grants” discussed by Bartel et al (2007). The recipient may feel a sense of belonging to the community that can in turn increase his or her willingness to participate (Blanchard and Markus 2004). Pronouns are especially powerful language features that signal the nature of the connection between speakers and audiences (Brewer and Gardner 1996, Brown and Gilman 1960, Pennebaker, et al. 2003). For example, first-person plural pronouns (“we”, “us”, “our”) are often used to signal solidarity between the writer and audience. Third-person pronouns (e.g., “him”, “them”) can have similar effects because they are often used to distinguish an in-group (“us”) from an out-group (“them”). In contrast, second-person pronouns (e.g., “you”) can be exclusionary, signaling a distinction between author and addressee (Brown and Gilman 1960) and suggesting that the responder does not consider the poster (“you”) to be part of the community. Although first-person plural pronouns can signal solidarity, when paired with second-person pronouns in a “we/you” formation their use may create a distance between speaker and audience. We thus expect that

H4: Social interaction will have a larger impact on individuals’ relationship to a community when a) it includes more inclusive language (i.e., “we” pronouns and “them” pronouns), and b) include less exclusionary language (i.e., “you” pronouns and “we/you” formations).

Relationship development from the community’s perspective

To this point, we have focused on one side of the relationship-development process — the factors that could strengthen a newcomer’s relationship with a community. But the individual-community relationship is bi-directional. While individuals are evaluating and interacting with the community, the community is also evaluating and responding to them. A community’s willingness to maintain a rela-

tionship with newcomers also depends on who the newcomers are and the social tones in newcomers' attempts to interact with the community.

The group socialization theory suggests that groups are more willing to engage individuals who have already shown commitment and who are willing and able to provide resources the group values (Moreland and Levine 1988, Moreland and Levine 2001). When an individual has made explicit interaction attempts in a community and contributed to conversations in the past, the community is more likely to deem the relationship with this individual mutually beneficial. Prior engagement behaviors may also increase the visibility and influence of an individual in a community, which makes it more attractive for other members to interact with them. Thus, a community may be less motivated to interact with newcomers who are less visible and have not contributed much in the past. Moreover, if an individual has participated extensively in the community, other community members may feel obliged to reciprocate by responding (Gouldner 1960, Wasko and Faraj 2005). Therefore a community is expected to be more willing to provide a response to individuals with a history of interaction with it.

H5: A community is more likely to maintain a relationship with individuals who have participated actively in the past.

The social characteristics of the messages that people post in an online community are also likely to influence whether the community will attempt to maintain a relationship with them. Just as the language that the community adopts in responding to individuals may influence the individuals' subsequent relationship with the community, the relational model of online communities predicts that the language that individuals use is likely to influence the community's willingness to engage with them. When posters signal a connection with a community through the use of inclusive language, other community members may feel more favorably about the posters and therefore be more likely to respond. Choices of words that indicate solidarity with the community, such as the pronoun "we" and "them," should increase the likelihood of the community responding. Not signaling an affiliation with the community, or using words that signal separation, such as like "you" and the "we/you" formation, may distance a poster from the community and thus reduce the willingness of the community to engage in interactions with them. Therefore

H6: A community is more likely to maintain a relationship with individuals whose posts a) use more inclusive language, and b) use less exclusive language.

Individuals may also use language to claim more directly that they have a relationship with the community. People can assert the existence of an individual-community relationship by noting that they are readers of the group's messages ("I've been lurking here a while and wanted to ask...") or by alluding to the group's shared history through reference to past posts or posters (Galegher, et al. 1998). They can also claim a relationship with the larger community from which the online community draws its members (e.g., "I was recently diagnosed with epilepsy," to an epilepsy support group). These efforts to linguistically establish an individual-community relationship occur in the type of message we call testimonials, messages that present a personal story indicating the individual's connection to the group either directly or by virtue of shared experience or common identity. Prior work has shown that testimonials are often included in individuals' first messages to a community (Burke, et al. 2007, Galegher, et al. 1998, Rafaeli, et al. 2004). These types of messages signal the authors' identification with the community and their willingness to develop a relationship with it. Therefore, an online community is expected to be more willing to engage individuals who include testimonials in conversation-initiation messages, because these messages signal the individuals' interest in maintaining a relationship with the community.

H7: A community is more likely to maintain a relationship with individuals whose posts include testimonials.

While testimonial language may be useful for newcomers to introduce themselves to a community, its effects should diminish as newcomers participate more. When newcomers first start to engage a community, their lack of prior exposure in the community makes it difficult for other members to evaluate them and thus makes others unwilling to respond. Testimonials provide a signal of intention that plays an important role in the community's decision to engage newcomers. As the community gets to know them through interaction, newcomers' testimonials should be less necessary and informative. Therefore, we expect that the effect of testimonial language should be greatest when individuals start to interact with a community and decrease as they interact with a community more.

H8: *Testimonials will have less impact on a community's maintaining a relationship with individuals the more they participate in the community.*

Methods

We tested the hypotheses derived from the relational model of online communities using a longitudinal dataset of 28,869 Usenet newcomers participating in 98 newsgroups over 650 days. Usenet is composed of over 190,000 online discussion groups, called newsgroups, involving millions of people discussing topics ranging from hobbies to technical support to politics. While there are other technologies that support communities, including mailing lists, web forums and social networking systems, text-based, threaded discussion forums remain a key element of many online communities (Kim 2000). Therefore, Usenet provides a rich environment for studying the dynamics of online communities (Smith, 2004).

Data

Our data comprise metadata and message text from 98 Usenet newsgroups from May 15, 2003 to February 23, 2005. The metadata, provided by Microsoft's Netscan project (Smith, 2005), includes information about groups, authors, and messages, such as the total number of messages posted to a group on a given day, dates of an individual's first and last posts to a group, and the number of replies that a message received. The text of posted messages was downloaded from a Usenet archive and matched to the metadata using a unique message identifier.

We randomly selected ninety-eight newsgroups that were active during the study period, stratified by broad topics. To ensure they were active, we excluded groups that had a mean of fewer than 3 messages per week during the study period. To reduce the chances that results would be topic- or population-dependent, we selected groups from four general categories: health support, technical support, hobby, and political issues. Health support groups deal with topics such as asthma, breast cancer, and food allergies; technical support groups with topics such as civil engineering, Windows NT security, and numeric analysis; hobby groups with topics such as aquarium plants, the Grateful Dead, and vegetarian cooking; political issue groups with topics such as gun rights, economics, and nationalism.

We selected a random sample of 28,869 unique individuals who posted for the first time in these groups between May 15, 2003, and Feb. 23, 2005. All the messages contributed by these individuals to the 98 newsgroups during the study period and the replies that followed were included in the dataset. To control for spam, a commercial spam filter, SpamAssassin, was used to calculate a spam score (ranging from 0 - 13) for each message. Messages with a spam score higher than 5, SpamAssassin's default spam level, were removed. Cross-posted messages sent to multiple newsgroups in the sample were dropped because they might involve competition and other dynamics peripheral to this study. The analysis dataset included 221,092 messages posted by 28,869 unique authors in 98 Usenet groups over 650 days.

Measures

Because all continuous variables in the dataset were highly skewed, a log (base 2) transformation was applied to them to increase the normality of the distribution. Table 1 shows descriptive statistics for variables in the dataset before the log transformation.

Table 1: Descriptive Statistics for Variables

Variable	N	Mean	Median	Min	Max	Std. Dev.
Individual duration in group (days)	28869	36.60	1	1	637	101.84
Future Reply Ratio	173717	0.88	0.97	0	1	0.24
Got Reply (0=no/1=yes)	221092	0.45	0	0	1	0.50
Number of replies received	221092	0.68	0	0	69	1.15
Poster's previous posts in newsgroup	221092	268.74	44	0	5136	586.39
Replier's previous posts in newsgroup (max)	99120	270.61	64	0	10155	561.52
IsTestimonial (0=no/1=yes)	221092	0.65	1	0	1	0.48
# singular first person pronouns in message	221092	3.24	1	0	692	7.98
# 1st person plural pronouns in message	221092	0.79	0	0	321	3.92
# 2nd person pronouns in message	221092	2.09	0	0	385	5.88
# 3rd person plural pronouns in message	221092	2.29	0	0	745	8.43
# 1st person plural pronouns in reply	99104	0.62	0	0	300	2.98
# 2nd person pronouns in reply	99104	2.14	1	0	361	5.77
# 3rd person plural pronouns in reply	99104	1.94	0.25	0	509	6.01
Day volume in newsgroup	221092	125.39	111	1	875	95.25
Word count in message	221092	125.78	51	0	15133	332.55
Words per sentence in message	221092	15.40	14	0	912	12.91
Word count in reply	99104	104.73	53	0	12741	237.37
Words per sentence in reply	99104	15.39	14.31	0	914.33	12.54

Dependent Variables

The strength of individuals' relationship with an online community is reflected in their engagement behaviors in the community over time. Their decision to continue to post messages and their willingness to engage in pro-social behavior that is helpful to other participants in the community are both indicators of their relationships with the community. People who continue to participate in a community over an extended period have a stronger relationship with it than those who participate only once or twice. In addition, individuals who reply to others' posts often have more group-oriented motives, i.e., more of a sense of relationship with the community, than those who start threads of their own (Bateman, et al. 2006). We measured two aspects of the strength of individuals' relationship with a community: duration of participation and pro-social contributions. Duration refers to the length of time an individual continues to participate in conversations in the community. It is measured with a set of dummy variables, *PostAgain*, from an individual's first day posting to his or her last post in a newsgroup. *PostAgain* takes the value of 1 if an individual posts another message after the current posting day, and 0 otherwise. Thus, an individual's *PostAgain* value will always be 1 until he or she stops participating in a newsgroup. Together, an individual's *PostAgain* values over time indicate the individual's length of participation in a newsgroup. Because individuals whose last post in a newsgroup appeared close to the end of data collection may have returned after the data collection, we treated those whose last post was within 15 days of the data collection end date as still remaining in the group. In other words, these individuals were right censored (Singer and Willett 2003).

FutureReplyRatio indicates the pro-social nature of participants' contributions. This variable measures the extent to which an individual's posts to the group were replies to others' messages rather than thread-starting messages, an indication that the individual is offering help instead of attempting to extract value from the group. The *FutureReplyRatio* is calculated as the number of replies divided by the total number of messages that an individual posts outside the current discussion thread in the next 30 days. We used a fixed 30-day window following a post to increase independence of the *FutureReplyRatio* observations for the same individual. This ratio could be calculated only for individuals who posted at least one other message in the following 30 days and were missing otherwise.

A community's relationship with an individual is reflected in the community's willingness to respond to the individual's attempt to start a conversation. Therefore, a community's relationship with its members is measured with a dummy variable *GotReply*, which indicates whether an initial message received a reply (1) or not (0).

Independent Variables

NumberOfReplies is used to measure the extent to which individuals receive responses from an online community, an indication that the community is interacting with them. It is calculated as the number of replies that an individual receives to a message.

We measured individuals' level of prior interaction with a community using the number of their previous posts to the community. We calculated an author's *PreviousPosts* as the total number of messages that an individual posted to a newsgroup before the current day. Similarly, for messages that received replies we calculated the *ReplierPreviousPosts* as the total number of messages that the replier posted to a newsgroup before the current day. When a message received multiple replies, *ReplierPreviousPosts* is the maximum of these values.

MessageInclusiveness indicates the degree to which a message uses welcoming or inclusive language. It is measured using the total count of first-person plural pronouns (e.g., "we"), second-person pronouns (e.g., "you") and third-person plural pronouns (e.g., "them") in a message, all calculated by the Text Analysis and Word Count (TAWC) program (Kramer, et al. 2004). The "we" and "them" pronouns are used as indications of inclusive language, and the "you" pronouns and "we/you" formulations indicates exclusionary language. These measures are constructed for both focal messages and replies. When a message received multiple replies, the count of pronoun measures for the reply messages is the average value among all replies.

We constructed two variables to measure the degree to which posters are claiming connection to a group with their posts: *IsTestimonial* and *FirstPersonSingularPronouns*. Because testimonials are used to imply a poster's already-existing relationship with the community, they often contain first-person pronouns, the age of the poster, a description of the poster's situation and history ("I was diagnosed in 1994

with BC...”), and a request for advice (“If I do sit-ups every day, say 15 or so, along with other exercises, how long until my stomach muscles show?”). We used the Minorthird machine learning and text classification toolkit (Cohen, 2004) to create an *IsTestimonial* measure, which is a dummy variable that takes the value 1 if the message was classified as a testimonial or 0 otherwise. To create the *IsTestimonial* variable, we first hand-coded 200 messages from two support groups not included in our analysis as being testimonials or not, and then used Minorthird to reproduce the hand classifications using the words in the original messages as the input. Based on a five-fold cross-validation of the 200 messages, agreement between the machine learning classification and the hand-coding was good (Kappa=0.79). We then applied the model developed from the 200 messages to the remainder of the dataset to create the *IsTestimonial* measure. Because self-introduction messages often involve the use of first person pronouns, we also measured the extent to which an individual signals connection to a group using *FirstPersonSingularPronouns*, the count of the number of first person pronouns (e.g. “I”, “me”, “my”) in a message based on the output of TAWC.

Controls

To account for specific group effects, 98 dummy variables were created to represent each newsgroup and were controlled for in the analyses. We also control for the information overload that individuals may face in newsgroups. Prior research has shown that high information processing costs associated with heavy message volumes and long messages can drive people away from online communities and reduce the likelihood that they will reply to the messages posted there (Butler 2001, Jones, et al. 2004). To control for information processing and attention costs, we included measures of message volume in a newsgroup, message length and message complexity in the analyses. Group message volume was measured with *NewsgroupDayVolume*, the total number of messages posted in the newsgroup on the same day as the focal message. *WordCount* is the number of words in a message. *WordsPerSentence*, calculated as the total number of words divided by the number of sentences in a message, indicates the complexity of the message. Word count and words per sentence were calculated both for the focal message and for reply

messages. If a focal message received multiple replies, these measures for the reply messages are calculated as the mean among all replies.

Analysis & Results

Modeling the duration of individuals' engagement with a community

To examine how social interaction with an online community influences the length of individuals' engagement with the community, we model the impacts of the number of replies individuals received, the prior experience of the repliers and the inclusiveness of the reply messages on individuals' participation duration in a newsgroup.

When the dependent variable involves time and possible censoring, survival analysis is the appropriate technique to use. Survival analysis is often used to study the time between entry to a study and the experience of an event (Singer and Willett 2003). Here it is applied to predict the time between an individual's first and last post in a newsgroup using features of the interaction between the individual and the community. Survival analysis requires data to be formatted so that there is one observation for each individual at each time period, with the dependent variable representing whether the individual experiences the event at the current time period or not. Because the smallest unit of time in our data is the day, we aggregate the message level of data at the individual-day level. The dependent variable *PostAgain* represents the occurrence or non-occurrence of the event, which is whether the individual has stopped participating in the newsgroup after the current day. If an individual posts multiple messages in a newsgroup on the same day, the characteristics of these messages and their replies are aggregated and the average values are used in the analysis. Therefore, while the survival analysis is based on all 221,092 messages that the 28,869 individuals posted across 650 days, the number of observations in this analysis is 105,455 individual-days. Figure 1 plots the survival graph, which shows the percentage of individuals remaining in a community over time, grouped by the number of replies they receive on a given day. The data have been aligned, so that zero on the x-axis indicates when a person first posted to a group. Among the 28,869 newcomers in the sample, the majority (69%) left after their first day, although some continued to participate in a newsgroup for up to 637 days. The separate curves show that people who receive more replies

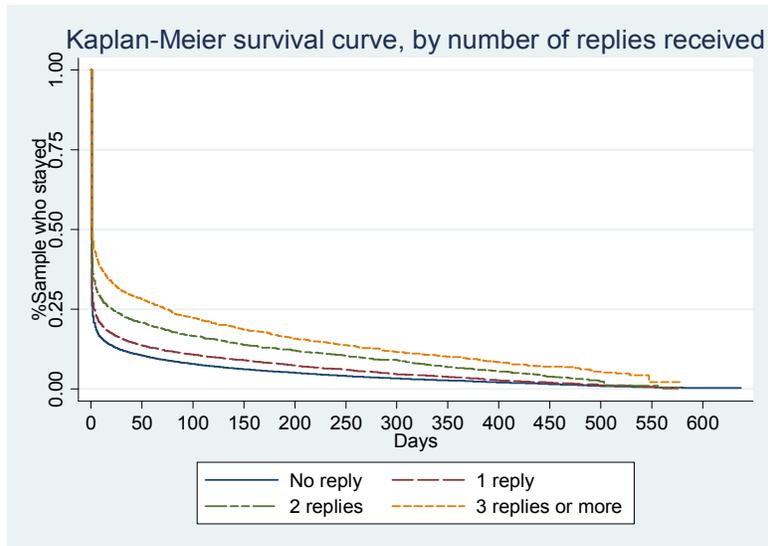


Figure 1: Percent of individuals remaining in sample over 600 days, by the number of replies received

stay in the group longer. By day 30, a month after their first appearance in a group, 31% of the people whose messages received three or more replies are still participating in the group, while only 12% of those whose messages received no replies still remain in the group. The gaps between those who got many replies and those who got none remained even a year after their first appearance in a group.

We conducted parametric survival analysis based on a proportional hazard model with a log-logistic distribution (Hosmer and Lemeshow 1999) to test hypotheses 1-4, using the `streg` procedure in Stata. VIFs for all the independent variables are below 5, indicating that multicollinearity is not a concern in this analysis. Table 2 reports the regression results. Model 1 in Table 2 shows that after accounting for the control variables, each doubling of the number of replies received increases the time an individual stayed in the community by 10.4%¹ ($\beta=.099$, $p<.001$), supporting Hypothesis 1. Model 2 adds the interaction between the number of replies and individuals' prior posting experience. Consistent with Hypothesis 2, the significant negative interaction between individuals' prior posts and the number of the replies that they received ($\beta=-.034$, $p<.001$) indicates that the impact of receiving replies decreases as individuals gain

¹ The hazard ratio is calculated from the coefficient: $e^{.099}=1.104$.

more experience in the newsgroup. These results show that while receiving replies from the community is always important, it is especially critical for newcomers with limited experience.

Model 3 in Table 2 shows that for individuals who received at least one reply for messages posted on a given day, the social status of the repliers and inclusiveness of the reply language had an impact on the length of their engagement with the community. Among the people who received replies, receiving replies from more active posters, i.e., those who posted many messages before, increased the length of their participation. Doubling the replier's prior posts increases an individual's duration with a group by 3.7% ($\beta=.036$, $p<.001$). Hypothesis 3 is thus supported. Receiving replies that use inclusive language also affect individuals' engagement with a community. Individuals participated in a group longer if they received replies containing more first-person plural pronouns (e.g., "we") ($\beta=.052$, $p<.01$), and for less time if replies contained more second-person pronouns (e.g. "you") ($\beta= -.044$, $p<.001$). Doubling the number of "we" pronouns increased time in the group by 5.3%, while doubling "you" pronouns reduced the time by 4.3%. Also, when the inclusive first-person pronouns and exclusive second-person pronouns were used together, the reply message signaled stronger separation between the community and the original poster, further reducing the person's likelihood of future engagement ($\beta=-.022$, $p<.001$). However, replies containing more third-person pronouns ("them") did not make a difference. Thus, Hypothesis 4 receives only partial support. Together the results from Model 3 suggest that the social status of repliers and the social tone of replies increase the strength of individuals' relationship with an online community as predicted by the relational model: receiving welcoming, inclusive responses from active members of the community encourages longer individual engagement above and beyond the effect of merely receiving responses.

Modeling the pro-social nature of individuals' community engagement

To look at another aspect of individuals' relationship with a community, we examine the impact of the same social features on individuals' tendency to exhibit pro-social behavior (*FutureReplyRatio*). Unlike the survival analysis, this analysis is conducted at the message level. Because messages posted by the same authors are not independent, mixed-model regression analysis was used to account for the nested

nature of the dataset, treating individual as a random effect. Not all individuals post another message after a focal message. A meaningful reply ratio can only be calculated for the 173,717 messages that are followed by at least one more post in the next 30 days. Therefore, we apply the Heckman two-stage procedure² to adjust for selection bias in the sample, taking into account the effect that independent variables may have on the likelihood that an individual will post another message in the next 30 days (Heckman 1979). VIFs for all independent variables are below 4, indicating that multicollinearity is not an issue.

Model 4 in Table 2 shows that the number of replies a message received had a small positive effect on the extent to which the poster's subsequent messages were replies ($\beta=.001$, $p<.01$), supporting Hypothesis 1. Doubling the number of replies a message received increased the proportion of replies the original poster offered in the next 30 days by 0.1%. Model 5 added the interaction between the number of replies and the number of the poster's prior posts in the community. Consistent with Hypothesis 2, the impact of receiving replies decreased as individuals interacted with the group more extensively ($\beta = -.0003$, $p<0.01$). Although the coefficients in these models are greater than zero, supporting Hypotheses 1 and 2, the effect sizes are very small and are unlikely to be of substantive interest.

Replier and reply message characteristics were added in Model 6, and their impacts were examined for the 76,470 messages that received a reply. Repliers' prior number of posts in the group did not affect an individual's future reply ratio, providing no support for Hypothesis 3. This suggests that an individual's tendency to behave pro-socially in the near future is not influenced by repliers' status. Replies containing more third-person pronouns (e.g. "them") increased the future reply ratio ($\beta = 0.0018$, $p<0.001$). However, none of the other language characteristics were significant, providing little support for Hypothesis 4.

² An inverse mills ratio is first calculated when modeling the likelihood of an individual posting another message in the next 30 days with the independent variables. Then, the inverse mills ratio is included in the regression predicting Future Reply Ratio to adjust for selection bias.

Table 2: Individual perspective: predicting duration and future reply ratio in newsgroup

Coefficient	Predicting Duration			Predicting Future Reply Ratio		
	Model 1: Main	Model 2: Interaction	Model 3: ReplyChar	Model 4: Main	Model 5: Interaction	Model 6: ReplyChar
Constant	4.376*** (0.076)	4.357*** (0.076)	4.508*** (0.118)	0.543*** (0.041)	0.543*** (0.041)	0.570*** (0.044)
Day volume in newsgroup	0.062*** (0.008)	0.062*** (0.008)	0.023 (0.013)	0.001* (0.000)	0.001* (0.000)	0.001 (0.001)
Word count in message	0.001 (0.004)	0.001 (0.004)	0.008 (0.007)	-0.001*** (0.000)	-0.001*** (0.000)	-0.001** (0.000)
Words per sentence in message	0.014* (0.006)	0.014* (0.006)	0.001 (0.009)	0.001*** (0.000)	0.001*** (0.000)	0.001** (0.000)
Poster's previous posts in newsgroup	0.879*** (0.007)	0.874*** (0.007)	0.891*** (0.013)	-0.003*** (0.000)	-0.003*** (0.000)	-0.004*** (0.000)
Number of replies received	0.099*** (0.008)	-0.028 (0.027)	-0.266*** (0.047)	0.001** (0.000)	0.001** (0.000)	0.000 (0.001)
Previous posts*Number of Replies		-0.034*** (0.007)	-0.092*** (0.012)		-0.000** (0.000)	0.000 (0.000)
Replier's previous posts in newsgroup			0.036*** (0.004)			-0.000 (0.000)
# 1st person plural pronouns in Reply (e.g. we)			0.052** (0.019)			0.001 (0.001)
# 2nd person pronouns in reply (e.g. you)			-0.044*** (0.011)			-0.000 (0.000)
# 3rd person plural pronouns in reply (e.g. them)			-0.008 (0.015)			0.002*** (0.001)
We*you			-0.022** (0.008)			0.000 (0.000)
Word count in reply			0.021* (0.01)			-0.001* (0.000)
Words per sentence in reply			0.006 (0.01)			0.000 (0.000)
Inverse mills ratio	- -	- -	- -	-0.021*** (0.002)	-0.021*** (0.002)	-0.038*** (0.004)
Observations	105455	105455	59320	173717	173717	76470
AIC	103749	103727.5	50918.44	-317482.6	-317468.9	-132792.4

Note: *** p<0.001, ** p<0.01, * p<0.05; Standard errors in parentheses.

* Results for the 97 newsgroup dummies are omitted for simplicity.

Community Perspective

To examine the factors that lead a community to engage newcomers, we analyzed how characteristics of the individuals and features of their attempted interactions affect the community's willingness to respond to them. Because the community's reaction to participants' attempts to initiate conversations is measured by a binary outcome (*GotReply*), probit analysis was used to examine the outcome for 30,508

initiating messages, i.e., messages that were not themselves replies to other messages. Messages posted by the same individual are not independent, and thus we used robust standard errors with clustering to adjust for the non-independency. Table 3 shows the results of probit regressions. The coefficients in Table 3 represent percentage changes of the probability that a message will receive a reply when the value of the independent variable doubles or, in the case of binary variables, changes from zero to one. VIFs for all independent variables were below 4.

Model 1 in Table 3 shows the impact of an individual's prior participation on the community's likelihood of providing a response. Hypothesis 5, which predicted that the group would be more responsive to those who have posted many messages in the past, was not supported. The coefficient for the individual's number of prior posts was negative and non-significant, suggesting that individuals' prior engagement in a community does not influence the community's willingness to respond.

Model 2 adds the characteristics of the initiating messages. Messages were less likely to receive a reply if they used exclusionary second-person 'you' pronouns ($\beta = -0.041$, $p < 0.001$) or the 'we/you' combination of first-person plural and second-person pronouns, which signal further differentiation between the poster and the community ($\beta = -0.007$, $p < 0.001$). However, the use of the first-person plural pronoun alone also reduced the likelihood of receiving a reply ($\beta = -0.016$, $p < 0.001$), suggesting that a poster's attempt to signal relationship with the community may not always be recognized by the community. The use of the third-person plural pronoun 'them' did not have an impact. These results show that while using exclusive language makes it harder to get a response from the community, using inclusive language in an initiating message may not be appropriate either. Therefore, Hypothesis 6 only received partial support. Hypothesis 7, which predicted that the group would be more responsive to those who introduced themselves and signaled their connection to the group through testimonials and first-person singular language, was supported. Including a testimonial increased the reply probability by about 9% ($\beta = 0.094$, $p < 0.001$). Doubling the use of first-person singular pronouns increased the reply probability by about 6% ($\beta = 0.058$, $p < 0.001$).

Model 3 added the interaction term between IsTestimonial and an individual's prior level of participation. The interaction between IsTestimonial and an individual's number of prior posts is negative but non-significant, which suggests that the impact of a testimonial in a thread-initiating message does not change after the individual posts more messages to the group. Thus, Hypothesis 8 is not supported.

Table 3: Community perspective: predicting GotReply

Coefficient	Predicting GotReply		
	Model 1: Sender	Model 2: Message	Model 3: Interaction
Day volume in newsgroup	0.026*** (0.007)	0.028*** (0.007)	0.027*** (0.007)
Word count in message	0.013* (0.007)	-0.006 (0.010)	-0.007 (0.010)
Word per sentence in message	0.022*** (0.006)	0.016*** (0.004)	0.016*** (0.004)
Poster's previous posts in newsgroup	-0.004 (0.004)	-0.001 (0.004)	-0.000 (0.004)
# 1st person plural pronouns in Message (e.g. we)		-0.019*** (0.006)	-0.016** (0.005)
# 2nd person pronouns in message (e.g. you)		-0.036*** (0.007)	-0.029*** (0.007)
# 3rd person plural pronouns in message (e.g. them)		0.007 (0.010)	0.009 (0.010)
We*you		-0.007*** (0.002)	-0.008*** (0.002)
IsTestimonial		0.094*** (0.014)	0.075** (0.023)
# 1st person singular pronouns in message (e.g. I)		0.058*** (0.004)	0.060*** (0.004)
Previous posts*Testimonial			-0.008 (0.006)
Observations (messages)	30508	30508	30508
Observed p	0.550	0.550	0.550
Predicted p	0.557	0.560	0.560
AIC	35679.8	34921.59	34893.54

Note: *** p<0.001, ** p<0.01, * p<0.05; Standard errors in parentheses.

* Results for the 97 newsgroup dummies are omitted for simplicity.

Our failure to find that groups were more responsive to individuals who posted frequently in the past was surprising. Therefore, we conducted post-hoc analysis to further examine the impact of different types of prior participation. When breaking prior posts into prior replies and prior initiating messages, the results show that groups were more likely to respond to people who had posted many replies in the past ($\beta = 0.027$, $p < 0.001$), and less likely to respond to those who started many threads ($\beta = -0.032$, $p < 0.001$).

This suggests that groups are sensitive not just to the extent to which an individual has interacted with the community, but also to the nature of that interaction. Moreover, there is a statistically significant interaction between the use of a testimonial and the number of replies an individual posted in the past ($\beta = -0.024$, $p < 0.001$), meaning that including testimonials is more effective in soliciting responses for people who have not posted replies before, but less useful for those who have replied to others.

Discussion

This paper supplements the information-centric view of online communities common in the information systems literature by proposing and testing a relational model of online communities. The relational model argues that information-sharing behaviors observed in online communities should be thought of as elements of a bi-directional, interaction-based social relationship developing between some individuals and a community (Figure 2). Overall, the empirical findings are consistent with the relational model. While almost 70% of the newcomers to Usenet groups posted only once, others developed longer-term relationships. Individuals' relationship with the community, measured in terms of the duration of their participation, is influenced by the number of replies they received from others in the community, the experience of the repliers and the inclusiveness of the language used in the replies. In addition, the impact of receiving replies on the strength of a longer-term relationship was greatest early in newcomers' history with the group and declined as they interacted more.

A relational model highlights the bi-directional nature of the relationship between individual and community. People can gain benefits from an online community only if the community is willing to engage them. Again, the findings are largely consistent with hypotheses from the relational model. Online communities were more likely to reply to individuals whose prior interactions with the community provided help to others and were less likely to reply to those who had posted many questions and other conversational starters. Communities were also sensitive to the social tone of the language that individuals used. Communities were more responsive to posters who used implicit membership claims and self-introductory testimonials signaling that they had an existing relationship with the community, and were less responsive when they used exclusionary language, such as "you" and "we/you" formulations that

separated them from the community. The impact of using testimonials on community response was greatest for newcomers with limited replying experience, and declined when they replied to others more frequently. Taken together, the results presented above support the idea that engagement in an online community can be thought of as deriving from an evolving, bi-directional, interaction-based social relationship between the individual and the community.

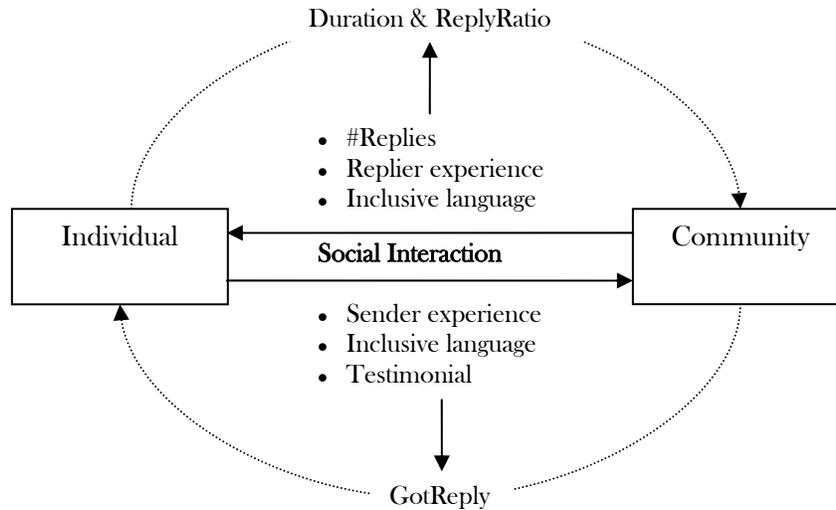


Figure 2: Relational Model of Online Communities

However, not all the results were consistent with the relational model. While individuals were more likely to remain engaged when responses they received used inclusive language, communities were *less* likely to respond to individuals who used the same “we” language. Post-hoc examination of a sample of messages and replies suggests that this result may partially derive from the differences in how we-pronouns are used. In responses, “we” typically refers to the online community that is responding (e.g., “I have no idea what you have in your code, seeing as we can only view your log file output.”). In contrast, when used by individuals in their initial messages, “we” frequently refers to outside groups that the individual is associated with rather than to the online community. Another possibility is that members of a community were put off when newcomers presumptuously use we-pronouns to claim group membership before they were granted it. Sorting out these competing explanations requires further research to examine

in more detail how the language used in online interaction affects the individual-community relationship, and how the framing of that relationship promotes or hinders its development.

Although the relational model does well at explaining why newcomers continue to participate in online communities for a long time, it does a poorer job of explaining the pro-social nature of that participation. Although the basic predictions — that individuals acted in a more pro-social way when their own posts received more replies, and that this effect was strongest earlier in their interaction with a community — were supported, the effects were too small to have practical importance. Furthermore, when people got replies to their initial posts, the repliers' social status and the inclusive language used in replies had little effect on subsequent pro-social behavior.

It is both surprising and interesting that the relationship model can explain the length of newcomers' engagement with the community but not the pro-social nature of the interaction. This difference is inconsistent with prior research in organizational studies, which has found that the strength of employees' relationships with a firm, measured by psychological commitment, is associated with both the duration and pro-social nature of their participation in the organization, assessed in terms of non-turnover and organizational citizenship behavior, respectively (Chen, et al. 1998, LePine, et al. 2002, Organ and Ryan 1995). However, in the current study, duration and pro-social behavior were only weakly related ($r=.09$) and were driven by different factors. It is possible that this result arises from the fact that the measures of duration and pro-social behavior are imperfect measures of strength of relationship; another possibility is that individual-community relationships, while they have some commonality with individual-organization relationships, also differ in significant ways. Additionally, the relationship between individuals and their communities may be multi-dimensional, with different dimensions affecting either individuals' survival or their replying behaviors, but not both. This speculation is consistent with empirical findings which show that distinct features of knowledge-contribution communities affect participation of different kinds, such as peripheral versus active participation (Koh, et al. 2007, Majchrzak, et al. 2006). Future research can build on the relational model of online communities by further examining the nuances associated with different aspects of an individual-community relationship and the various engagement behaviors.

Limitations

As with any study, this work has limitations that should be taken into account when interpreting the results. Our results show that social language can have a significant impact on individual and community willingness to continue to engage in interactions. The pronoun measures, however, were assessed with methods that rely on relatively primitive word counting approaches to language. These tools rely on the naïve assumption that the meaning and function of text arise simply from the sum of its words. By contrast, the machine learning approaches used in this study to identify testimonials incorporate aspects of the text overlooked by the more primitive tools, offering alternative strategies for analyzing language. Future research should work toward refining these tools and measures to allow scholars to develop and test richer theories regarding the role of language in online social interaction.

As is the case in many studies of online communities, the data examined here are limited to public behavior. The results presented above show that the relational model of online communities can explain some the public behavior that reflects the relationship between individuals and a community. However, individuals' behavior before their first posts and after their last, their private communication and their private beliefs are invisible. Private beliefs and interactions may operate differently than public, active interaction, so that these 'invisible' behaviors and beliefs could be a source of noise in the data that obscured the results. For example, although our data show that self-introductory testimonials seem to elicit responses from existing members of a community, the public nature of the data prevents us from knowing whether those who post testimonials differ on some invisible lurking behavior from those who do not. More fundamentally, the existence of other forms of engagement suggests interesting questions for future research regarding the development and impact of the individual-community relationship. For example, do social language and the status of responders affect peripheral observers in the same way that they affect active participants? The relational model of online communities suggests that it should, but not as much (because social language remains a signal, albeit a weaker one, of community intent and expectations). By considering this type of question, future research based on the relational model can move toward more complete and coherent explanations of the range of behavior seen in online communities.

Lastly, while the analysis is based on a topically diverse sample of communities, they share a common infrastructure—Usenet—and an old one at that. Further research could better understand the role of specific technologies in shaping community and individual engagement by studying online communities that use technologies such as listservs, web forums, social networking sites, blogs or multiplayer gaming platforms.

Contribution & Implication for Research

The relational model of online communities extends the information-centric and resource-based models of online communities common in the IS research literature. In a resource-based view, the central research question is explaining why people contribute valuable resources, such as time, information and attention, to a community (Butler 2001). Prior studies have identified different motivations for contributing that revolve around the benefits derived from participation, such as getting answers to questions, developing an expert reputation or improving one's earnings (Hars and Qu 2002, Hertel, et al. 2003, Roberts, et al. 2006). The relational model of online communities acknowledges the role that receiving benefits plays in shaping relationships with an online community. However, these instrumental benefits are only part of the story. By drawing on models articulated by group and organizational scholars, this work extends the IS literature on online communities by focusing attention on the social relationships that develop between the individual and the community, which can be valued in their own right, and the bi-directional, evolving, interaction-based way that they develop. At the same time, it contributes to organizational and group studies by using the online context to examine some of the communication actions through which the social relationships develop and group socialization occurs.

Although the relational model of online communities draws from the social psychological literature, it is not a complete theory of how psychological commitment develops between an individual and an online community. Rather it is a theoretical framework through which to view online communities, one that treats online communities as full-fledged social groups in which relationships between the members and the communities are developed and maintained over time through interaction. Although we believe that individuals are developing commitment towards communities and vice versa, we have no direct evi-

dence, such as attitudinal self-report measures, for these psychological processes. However, the results presented here suggest that there is explanatory value to the concept of an individual-community relationship, whether or not it has any psychological validity with the individuals involved. Across a range of communities, individuals respond to social language and other aspects of the interaction as if they were responding to welcomes from the community. Across a large number of interactions, communities respond to signals that newcomers have invested effort in the community.

At the same time, the relational model of online communities highlights important questions about the nature of individual-group relationships. For example, social psychologists suggest that individuals are attached to groups for different reasons. Some form a relationship with a group through their friendships with particular people in it, and others feel connected to the group as a whole because of their identification with the social category or topics the group represents (Sassenberg 2002). If people form a relationship with an online community in part by developing interpersonal ties with other people who participate in it, the rich theoretical literature in social psychology on the formation and consequences of interpersonal relationships, with its emphasis on attraction, exchange, and self-disclosure (Berscheid and Reis 1998) can be applied to online communities. If people's relationship to the online community is primarily based on the identity of the group, the large body of work on social identity and social categorization can offer a great deal on understanding issues such as member categorization, group formation and in-group and out-group conflicts in online communities (Hogg and Terry 2000). Findings of this study and this literature imply that scholars interested in explaining individuals' behavior within online communities should consider not only interpersonal links, but also the evolving, bi-directional relationship between the individual and the community as a whole.

The relational model can also help to understand information systems as embedded in larger social systems. While it might be tempting to think of online communities as mere tools, systems that provide benefits for users, the results presented above suggest that conceptualizing online communities in terms of 'use' or 'resources' fails to account for significant aspects of the behavior in them. Although a relational model is especially applicable to online communities, because they are fundamentally social

systems, this model could also be applied more broadly as an alternative perspective on traditional questions of technology use and acceptance. Even conventional information systems, such as data warehouses or Customer Relationship Management systems, involve many of the characteristics of social entities with diverse groups of people (designers, users, managers, administrators) using information technology as a platform for sharing information, interacting, expressing intent, and conveying expectations. Studies of technology anthropomorphism and socially situated technology use suggest that under many conditions individuals treat information systems as social entities and form relationships with them (Nass 2000). Future work not just on online communities, but also with a wider range of systems, should consider the potential value in conceptualizing the use of information systems as a bidirectional, interaction-based relationship as opposed to a unidirectional choice. For instance, perhaps the user-friendliness that designers build into systems is the analog of the welcoming behavior exhibited in online communities.

Implications for Practice

Thinking of online communities through a relationship lens also has implications for the infrastructure and policies that online community leaders should put into place to foster the relationships they desire between the individual and the community. The relational model of online communities provides an explanation for why communities are often subject to high turnover and short involvement lifespan. Newcomers initially investigating the community are likely to leave if they do not feel welcomed. They are most sensitive to the presence or absence of social aspects of interaction with the community when they first start to interact with it. At the same time, newcomers who are not aware of the appropriate ways to signal their intentions are least likely to receive responses from a community. Our results imply that there are many opportunities for a “failed relationship,” especially early in the relationship development process. Therefore, it is sensible that communities seeking new participants should institutionalize welcoming behavior.

The community developing Wikipedia recognizes the importance for building long-term relationships of a welcoming stance towards newcomers, and they implement this in the explicit policy known as "Please do not bite the newcomer," which states that “New contributors are prospective ‘members’ and

are therefore our most valuable resource. We must treat newcomers with kindness and patience — nothing scares potentially valuable contributors away faster than hostility.”³ In addition to admonishing its members against acting harshly towards newcomers, Wikipedia has instituted a welcoming committee whose main activity is to greet new users by posting a welcome message on the newcomers’ talk page, a place for interactions among members of Wikipedia. Members of the welcoming committee are encouraged to post a personalized greeting written by the committee members on the new user’s talk page.⁴ These policies and structures all reflect the recognition that developing the individual-community relationship is an integral part of building the community.

Although Wikipedia formally organizes people to welcome newcomers, many other online communities do not. For example, most open source development projects on Sourceforge.com do not have a mechanism in place to solicit and welcome potentially useful new members, and this perhaps explains why 75% of them never get a new participant after listing their project (Hahn, et al. In press). It is far more common for communities to begin interactions with individuals through Frequently Asked Questions (FAQs) that provide the group’s mission and rules of conduct, or for automatic agents to send newcomers boilerplate welcome messages. Although these techniques may serve to convey useful information, the relational model of online communities suggests that they will have minimal impact on engagement because they do little to support the process of building the individual-community relationship. At a minimum, introductory documents should be written with a focus on relationship-building. Beyond that, technological interventions such as computational agents which identify initiating posts and intervene if others are not responding within a reasonable time frame are likely to further strengthen the bonds between a community and the individuals involved. For example, if a newcomer asks a question about a drug in a cancer community but receives no response, an agent might forward the post to someone who has been active recently and has posted messages about drug therapies.

At the same time, individuals need to learn how to interact effectively with a community. While

³ http://en.wikipedia.org/wiki/Wikipedia:Please_do_not_bite_the_newcomers

⁴ http://en.wikipedia.org/wiki/Wikipedia:Welcoming_committee

some communities provide guidelines for participants, they focus on social control and etiquette, rather than fostering the behaviors and language that contribute to development of the individual-community relationship. Beyond this, the machine learning technique used in this study might also be used to analyze newcomers' initiating posts and provide automated interventions, such as suggesting wording changes, providing tips on similar subjects or identifying successful examples, all of which could help newcomers interact more effectively with the community.

Conclusion

This study proposes and empirically tests a relational model of online communities. We show that the social characteristics of interactions between individuals and the larger community reinforce the development and maintenance of the social relationship between them. The findings and arguments presented here identify next steps in explaining and improving an important class of online behavior. In addition to providing a foundation for understanding how people engage communities, they also challenge IS researchers to revisit their assumptions about systems in general and how people engage them. While a relational approach may not apply in all cases, as systems become more adaptive, dynamic, and social, and as the range of groups that information systems professionals are called upon to support grows, the explanatory power and applicability of the relational model is likely to increase.

References

- Ahuja, M. K. and Galvin, J. E., "Socialization in virtual groups," *Journal of Management*, 29, 2, (2003), 161-185.
- Arguello, J., Butler, B. S., Joyce, L., Kraut, R., Ling, K. S., Rosé, C. P. and Wang, X., "Talk to me: Foundations for successful individual-group interactions in online communities," In *CHI 2006: Proceedings of the ACM conference on human-factors in computing systems*, ACM Press, New York, 2006, 959 - 968.
- Bartel, C. and Dutton, J., "Ambiguous Organizational Membership: Constructing Organizational Identities in Interactions with Others," In *Social Identity Theory and Organizational Processes*, M. A. Hogg and D. J. Terry (Ed.), Psychology Press, London, 2001, 115-130.
- Bartel, C. A., Wrzesniewski, A. and Wiesenfeld, B., "The Struggle to Establish Organizational Membership: Newcomer Socialization in Remote Work Contexts," In *Identity and the Modern Organization*, C. A. Bartel, S. Blader and A. Wrzesniewski (Ed.), Lawrence Erlbaum, Mahwah, NJ, 2007, 119-134.
- Bateman, P., Butler, B. and Gray, P., "Community Commitment: How Affect, Obligation, and Necessity Drive Online Behaviors," *Proceedings of the Twenty Seventh International Conference on Information Systems, Milwaukee 2006*, 2006,
- Bauer, T. N., Bodner, T., Erdogan, B., Truxillo, D. M. and Tucker, J. S., "Newcomer Adjustment During Organizational Socialization: A Meta-Analytic Review of Antecedents, Outcomes, and Methods," *Journal of Applied Psychology*, 92, 3, (2007), 2707-2721.
- Berger, J., Fisek, M. H., Norman, R. Z. and Zelditch, M., Jr., *Status characteristics and social interaction: An expectation states approach*, Elsevier, New York, 1977.
- Berscheid, E. and Reis, H. T., "Attraction and close relationships," In *The handbook of social psychology*, D. T. Gilbert and S. T. Fiske (Ed.), 2, McGraw-Hill, New York, NY, US, 1998, 193-281.
- Berscheid, E., Snyder, M. and Omoto, A., "Conceptualizing and measuring closeness," In *Issues in studying close relationships* C. Hendrick (Ed.), Sage Publications, Newbury Park, CA, 1989, 63-91.
- Blanchard, A. L. and Markus, M. L., "The experienced "sense" of a virtual community: characteristics and processes," *ACM SIGMIS Database*, 35, 1, (2004), 64-79.
- Bock, G.-W., Zmud, R. W., Kim, Y.-G. and Lee, J.-N., "Behavioral Intention Formation in Knowledge Sharing: Examining the Roles of Extrinsic Motivators, Social-Psychological Forces, and Organizational Climate," *MIS Quarterly*, 29, 1, (2005), 87-111.
- Brewer, M. B. and Gardner, W., "Who is this "We"? Levels of collective identity and self representations," *Journal of Personality & Social Psychology*, 71, 1, (1996), 83-93.
- Brown, R. and Gilman, A. F., *The Pronouns of Power and Solidarity*, Bobbs-Merrill 1960.
- Burke, M., Joyce, E., Kim, T., Anand, V. and Kraut, R., "Introductions and Requests: Rhetorical Strategies That Elicit Online Community Response," *Proceedings of the The third communities and technologies conference*, New York, NY, 2007, 21-40.
- Butler, B., "Membership size, communication activity, and sustainability: A resource-based model of online social structures," *Information Systems Research*, 12, 4, (2001), 346-362.
- Butler, B. S., "When is a Group not a Group: An Empirical Examination of Metaphors for Online Social Structure," New Orleans, LA., (2004),
- Chen, X. P., Hui, C. and Seago, D. J., "The role of organizational citizenship behavior in turnover: Conceptualization and preliminary tests of key hypotheses," *Journal of applied psychology*, 83, 6, (1998), 922-931.

- Collins, D. R. and Stukas, A. A., "The Effects of Feedback Self-Consistency, Therapist Status, and Attitude Toward Therapy on Reaction to Personality Feedback," *Journal of Social Psychology*, 146, 4, (2006), 463-483.
- Constant, D., Sproull, L. and Kiesler, S., "The kindness of strangers: The usefulness of electronic weak ties for technical advice," *Organization Science*, 7, 2, (1996), 119-135.
- Davison, K. P., Pennebaker, J. W. and Dickerson, S. S., "Who talks? The social psychology of illness support groups," *American Psychologist*, 55, 2, (2000), 205-217.
- Dellarocas, C., "Strategic Manipulation of Internet Opinion Forums: Implications for Consumers and Firms," *Management Science*, 52, 10, (2006), 1577-1593.
- Ducheneaut, N., "Socialization in an Open Source Software Community: A Socio-Technical Analysis," *Computer Supported Cooperative Work*, 14, 4, (2005), 323 - 368
- Ducheneaut, N., Yee, N., Nickell, E. and Moore, R. J., "The life and death of online gaming communities: A look at guilds in world of warcraft," *Proceedings of the SIGCHI conference on Human factors in computing systems*, San Jose, California, USA, 2007, 839 - 848.
- Duck, S. W., *Relating to others*, Dorsey, Chicago, 1988.
- Duck, S. W., *Human relationships*, Sage Publications, Thousand Oaks, CA, 1998.
- Duck, S. W., Rutt, D. J., Hurst, M. H. and Strejc, H., "Some Evident Truths About Conversations in Everyday Relationships," *Human Communication Research*, 18, 2, (1991), 228-267.
- Eisenberg, E. M., Monge, P. R. and Miller, K. I., "Involvement in communication networks as a predictor of organizational commitment," *Human Communication Research*, 10, 2, (1983), 179-201.
- Galegher, J., Sproull, L. and Kiesler, S., "Legitimacy, authority, and community in electronic support groups," *Written Communication*, 15, 4, (1998), 493-530.
- Golder, S. A. and Donath, J., "Social roles in electronic communities," *Internet Research*, 5, (2004), 19-22.
- Gouldner, A. W., "The Norm of Reciprocity," *American Sociological Review*, 25, 2, (1960), 161-178.
- Granovetter, M. S., "The Strength of Weak Ties," *The American Journal of Sociology*, 78, 6, (1973), 1360-1380.
- Gu, B., Konana, P., Rajagopalan, B. and Chen, H. W. M., "Competition Among Virtual Communities and User Valuation: The Case of Investing-Related Communities," *Information Systems Research*, 18, 1, (2007), 68-85.
- Hahn, J., Moon, J. Y. and Zhang, C., "Emergence of new project teams from open source software developer networks: Impact of prior collaboration ties," *Information Systems Research*, (In press),
- Hars, A. and Qu, S., "Working for Free? Motivations for Participating in Open-Source Projects," *International Journal of Electronic Commerce*, 6, 3, (2002), 25-39.
- Heckman, J., "Sample selection bias as a specification error," *Econometrica*, 47, (1979), 153-161.
- Hertel, G., Niedner, S. and Herrmann, S., "Motivation of software developers in Open Source projects: an Internet-based survey of contributors to the Linux kernel," *Research Policy*, 32, 7, (2003), 1159-1177.
- Hill, K. and Hughes, J., *Cyberpolitics: Citizen Activism in the Age of the Internet*, Rowman & Littlefield Publishers, Inc. , Lanham, MD, 1998.
- Hogg, M. A. and Terry, D. J., "Social identity and self-categorization processes in organizational context," *Academy of Management Review*, 25, 1, (2000), 121-140.
- Homans, G. C., "Social behavior as exchange," *The American Journal of Sociology*, 63, 6, (1958), 597-606.

- Hosmer, D. W. J. and Lemeshow, S., *Applied Survival Analysis*, John Wiley & Sons, Inc, 1999.
- Hovland, C. I. and Weiss, W., "The influence of source credibility on communication effectiveness," *Public Opinion Quarterly*, 15, (1952), 355-350.
- Jones, G. R., "Socialization tactics, self-efficacy, and newcomers' adjustments to organizations," *The Academy of Management Journal*, 29, 2, (1986), 262-279.
- Jones, Q., Ravid, G. and Rafaeli, S., "Information overload and the message dynamics of online interaction spaces," *Information Systems Research*, 15, 2, (2004), 194-210.
- Kankanhalli, A., Tan, B. C. Y. and Kwok-Kee, W., "Contributing Knowledge to Electronic Knowledge Repositories: An Empirical Investigation," *MIS Quarterly*, 29, 1, (2005), 113-143.
- Koh, J., Kim, Y., Butler, B. and Bock, G., "Encouraging participation in virtual communities," *Communications of the ACM*, 50, 2, (2007), 68-73.
- Kramer, A. D. I., Fussell, S. R. and Setlock, L. D., "Text analysis as a tool for analyzing conversation in online support groups," *Extended Abstracts of the 2004 conference on Human Factors and Computing Systems (CHI 2004)*, (2004),
- Kristof, A. L., "Person-organization fit: An integrative review of its conceptualizations, measurement, and implications," *Personnel Psychology*, 49, 1, (1996), 1-49.
- Lakhani, K. R. and Hippel, E. V., "How open source software works: "Free" user to user assistance," *Research Policy*, 32, (2003), 923-943.
- Lakhani, K. R. and Von Hippel, E., "How open source software works: "Free" user to user assistance," *Research Policy*, 32, (2003), 923-943.
- LePine, J. A., Erez, A. and Johnson, D. E., "The nature and dimensionality of organizational citizenship behavior: A critical review and meta-analysis," *Journal of Applied Psychology*, 87, 1, (2002), 52-65.
- Liedtke, M., "Microsoft deal values facebook at \$15b," Thursday October 25, (2007),
- Lyytinen, K., "Implications of Theories of Language for Information Systems," *MIS Quarterly*, 9, 1, (1985), 61-74.
- Ma, M. and Agarwal, R., "Through a Glass Darkly: Information Technology Design, Identity Verification, and Knowledge Contribution in Online Communities," *Information Systems Research*, 18, 1, (2007), 42-67.
- Miller, V. D. and Jablin, F. M., "Information seeking during organizational entry: Influences, tactics, and a model of the process," *The Academy of Management Review*, 16, 1, (1991), 92-120.
- Moreland, R. L. and Levine, J. M., "Group dynamics over time: Development and socialization in small groups," In *The social psychology of time: New perspectives*, J. E. McGrath (Ed.), 91, Sage Publications, Thousand Oaks, CA, US, 1988,
- Moreland, R. L. and Levine, J. M., "Socialization in organizations and work groups," In *Groups at work: Theory and research*, M. E. Turner (Ed.), Lawrence Erlbaum Associates Publishers, Mahwah, NJ, US, 2001, 69-112.
- Mowday, R. T., "The Measurement of Organizational Commitment," *Journal of Vocational Behavior*, 14, 2, (1979), 224-247.
- Nass, C., "How People Treat Interfaces Like People: Social Psychology and Design," Seattle, WA, (2000),
- Organ, D. W. and Ryan, K., "A meta-analytic review of attitudinal and dispositional predictors of organizational citizenship behavior," *Personnel Psychology*, 48, 4, (1995), 775-802.
- Orlikowski, W. J. and Iacono, C. S., "Research Commentary: Desperately Seeking the "IT" in IT Research—A Call to Theorizing the IT Artifact," *Information Systems Research*, 12, 2, (2001), 121-134.

- Pennebaker, J. W., Mehl, M. R. and Niederhoffer, K. G., "Psychological aspects of natural language use: Our words, our selves," *Annual Review of Psychology*, 54, (2003), 547-577.
- Rafaeli, S., "Interactivity: From new media to communication," In *Sage Annual Review of Communication Research: Advancing Communication Science* 16, Sage, Beverly Hills, CA., 1988, 10-134.
- Rafaeli, S., Ravid, G. and Soroka, V., "De-lurking in virtual communities: a social communication network approach to measuring the effects of social and cultural capital," *Proceedings of the 37th Annual Hawaii International Conference on System Sciences*, Waikoloa, Big Island, Hawaii, 2004, 10.
- Ridings, C. M. and Gefen, D., "Virtual community attraction: Why people hang out online," *Journal of Computer Mediated Communication*, 10, 1, (2004), np.
- Roberts, J., Hann, I.-H. and Slaughter, S., "Understanding the Motivations, Participation and Performance of Open Source Software Developers: A Longitudinal Study of the Apache Projects," *Management Science*, 52, 7, (2006), 984 - 999.
- Sassenberg, K., "Common bond and common identity groups on the Internet: Attachment and normative behavior in on-topic and off-topic chats," *Group Dynamics: Theory and Practice*, 6, 1, (2002), 27-37.
- Saunders, C. S., Robey, D. and Vaverek, K. A., "The persistence of status differentials in computer conferencing," *Human Communication Research*, 20, 4, (1994), 443-472.
- Singer, J. D. and Willett, J. B., *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*, Oxford University Press, 2003.
- Sproull, L. and Kiesler, S. B., *Connections: New ways of working in the networked organization*, The MIT Press, Cambridge, MA, US, 1991.
- Stewart, K. J. and Gosain, S., "The Impact of Ideology on Effectiveness in Open Source Software Development Teams," *MIS Quarterly*, 30, 2, (2006), 291-314.
- Tyler, J. R., Wilkinson, D. M. and Huberman, B. A., "Email as Spectroscopy: Automated Discovery of Community Structure within Organizations.," *The Information Society*, 21, 2, (2005), 143-152.
- Van Maanen, J. and Schein, E. H., "Toward a theory of organizational socialization.," *Research in Organizational Behavior*, 1, (1979), 209-264.
- Venkatesh, V., Morris, M. G., Davis, G. B. and Davis, F. D., "User Acceptance of Information Technology: Toward a Unified View," *MIS Quarterly*, 27, 3, (2003), 425-478.
- Verona, G., Prandelli, E. and Sawhney, M., "Innovation and Virtual Environments: Towards Virtual Knowledge Brokers " *Organization Studies*, 27, 6, (2006), 765-788.
- Wagner, M., "Using Second Life As A Business-To-Business Tool," April 26, 2007, (2007),
- Wasko, M. M. and Faraj, S., "Why Should I Share? Examining Social Capital and Knowledge Contribution in Electronic Networks of Practice. ," *MIS Quarterly* , 29, 1, (2005),
- Weisband, S. P., Schneider, S. K. and Connolly, T., "Computer-Mediated Communication and Social Information," *Academy of Management Journal*, 38, 4, (1995), 1124-1151.
- Wellman, B., "Computer networks as social networks," *Science*, 293, 14 September, (2001), 2031-2034.
- Williams, D., Ducheneaut, N., Xiong, L., Zhang, Y., Yee, N. and Nickell, E., "From tree house to barracks: The social life of guilds in World of Warcraft," *Games and Culture*, 1, 4, (2006), 338-361
- Williams, K. D., Cheung, C. K. and Choi, W., "Cyberostracism: effects of being ignored over the Internet," *Journal of Personality & Social Psychology*, 79, 5, (2000), 748-762.