December 2003

Patterns of Chatter: An Empirical Case Study of Participation in an Online Health Community

Catherine Ridings
Lehigh University

PATTERNS OF CHATTER: AN EMPIRICAL CASE STUDY OF PARTICIPATION IN AN ONLINE HEALTH COMMUNITY

Catherine M. Ridings
College of Business and Economics
Lehigh University
Bethlehem, PA USA
ridings@lehigh.edu

Abstract

This research-in-progress is an in-depth case study of the patterns of interactions between participants in a virtual community. The study will examine the stability of the community membership, posting behavior of the members over time, the existence of clusters or groups of users, especially a core group, and the possible classification of members based upon participation. Social network analysis diagrams showing who is talking to whom will be produced, revealing the social structure of the community. All 16,112 messages posted by 1,670 users of a medical virtual community were gathered over the course of one year. Preliminary results are given and future analysis of the data is proposed.

Keywords: Virtual communities, social network analysis

Introduction

The Internet is increasingly a place to connect with others, especially in groups rather than one-on-one. Although large group conversations such as newsgroups on Usenet have been online for decades, it is only within the last 5 years or so that places such as chat rooms and bulletin boards have begun to multiply and attract mainstream Web users. An estimated 90 million Americans have used the Internet to contact an online group (79 percent remaining in contact with a particular group), and the majority of these individuals have been online 3 years or less (Horrigan et al. 2001). These persistent groups that arise as a natural consequence of people coming together online to discuss a common hobby, medical affliction, or personal experience, or even to develop relationships, have been termed virtual communities.

As the study of virtual communities continues to mature and expand, it is beneficial to seek a deeper understanding of the communication and membership patterns by identifying social connections in the community. Specifically examining the pattern of communication between participants can enlighten researchers and practitioners regarding the subgroups in the community, the stability of membership, and distribution of contributions by the members. Since the content of the community is developed by the members (rather than by the site owners), it is essential to understand who the contributors to the community are and their behaviors. This research will examine these issues by investigating the following research questions:

RQ1: Is the membership of a virtual community stable or does it change significantly over time?
RQ2: Do members start posting slowly and build up participation slowly over time?
RQ3: Do the activity levels reveal some kind of membership tiers? That is, can we classify users as low, medium, and high based on posting frequency?
RQ4: Do people cluster in groups—that is, do subgroups form in the community—and how do these clusters change over time?
RQ5: Is there a "core group" of posters that contribute more postings and have tight connections to one another?

Literature Review

Traditionally, communities have revolved around a common geographical location. Virtual communities expand this traditional meaning to instead view the communities as social networks or social relationships, typically centering on a common interest instead of shared geography. Howard Rheingold, in his famous book *The Virtual Community* (1993), defines the book's namesake as "social aggregations that emerge from the Net when enough people carry on those public discussions long enough, with sufficient human feeling, to form webs of personal relationships in cyberspace" (p. 5). Similarly, this research uses a definition from recent virtual community research: "groups of people with common interests and practices that communicate regularly and for some duration in an organized way over the Internet through a common location or mechanism" (Ridings et al. 2002, p. 273). The common virtual location, although not physical, is important because it establishes the virtual *place* where the members meet. This location or mechanism may be a chat room, bulletin board, or listserv program. Not all virtual meeting locations may be considered virtual communities. It is important that they have the defining elements of a notion of membership, relationships with others, commitment, distinctive focus, and existence for some duration (Erickson 1997; Figallo 1998) and minimum traffic and participation considerations (Ridings et al. 2002).

Virtual communities are fundamentally different than other groups that also use CMC, such as small work groups or groups formed to work on projects in classroom instruction. Virtual communities form of their own volition and their use is not mandated by organizations or instructors. They usually do not have a traditional leader or a specific work goal to accomplish. Wellman (1997) points out that virtual communities can be studied as either groups or social networks, defining a social network as "a set of people (or organizations or other social entities) connected by a set of socially meaningful relationships" (p. 179). Researchers have suggested the use of social network analysis to study computer-mediation communication (CMC) (Garton et al. 1997; Wellman 1996, 1997; Wellman et al. 1996). The present research specifically studies connections between community members using social network analysis. Social network analysis typically uses the relation or links between people as the fundamental unit of analysis: the set of relations is the social network (Garton et al. 1997). This study examines one attribute of relations, the strength of the relation, defined as the frequency of communication between people in the social network.

The essence of a virtual community is the conversation between the members. Thus the connections between members are an interesting area of study. Since the members themselves are the primary source of the content in the virtual community, it is important to understand their interactions in building the community knowledge. Involvement in the relationships between people in the community can help to explain attitudes and behaviors of the members and how the conversation is formed. For example, members who have many connections to other members may have higher satisfaction with the community. This knowledge can allow the community sponsor to take action to foster more participation, ultimately resulting in the success of the community. It has also been suggested the members can better use the community when they are aware of others' participation histories and relationships (Smith 2002). However, it is first necessary to determine the structure of the community before linking structural conditions with outcome variables such as satisfaction and community success. This research is focused on discovering the structural relationships within a particular virtual community.

Researchers have applied social network analysis to study the interaction of people via computers by examining the patterns of communication, flow of information, and direction and strength of the ties between people (Garton et al. 1997). Typically, social network analysis studies relations from either the perspective of a focal individual (ego-centered networks) or considering the whole network based on specific population criteria (Garton et al. 1997). Often the social network is drawn graphically with the nodes of the network representing people and the links indicating communication between the people. The length of the link may be shorter between people who communicate frequently, and longer to indicate infrequent communication. Several characteristics can be studied in social network analysis.

Membership Roles. Roles may be suggested by similarities in the behaviors. Some individuals may always be the first to respond to questions, and therefore adopt a teacher or "answer person" role. One way to categorize participants is as either core or periphery members by their frequency of posting as well as connections to others. Core members would post more frequently and have more connections to others. The size of a core group may indicate a community's maturity and stability (Smith 2002). Rice (1982) derived four types of members based on the flow of messages: isolates (sending and receiving less than average), receivers (receiving more than sending), transmitters (sending more than receiving), and carriers (sending and receiving more than average). The carrier type is analogous to the core group. Korenman and Wyatt's (1996) empirical study of a virtual community revealed a core of participants that contributed approximately 18 percent of the interaction in any given period.

Centrality. The degree of connectedness (either central or isolated) is known as centrality. Network members who have the most connections to others are said to have a high degree of centrality, and can be defined as a core group.

Size. Size of a community has been empirically linked to both member attraction (positively) and retention (negatively), but it is thought that information technology can reduce the negative effects of size (Butler 2001). Certainly larger communities have more resources available for members and it is more likely that another person will have the knowledge and time to answer a member's posting (Butler 2001). However, very large communities have the problem of information overload with many messages generated (Jones and Rafaeli 1999) and the increasing likelihood of logistical problems with getting to know the other members and forming personal relationships (Butler 2001). When communities become too large, the sense of community can be lost (Hiltz and Wellman 1997; Korenman and Wyatt 1996).

Network Density. In a dense network, most possible ties exist. In a sparse network, individuals rarely communicate directly and frequently with each other. Some virtual communities will have networks that are sparsely knit, with only a minority of members directly connected with each other, while others are tightly bound, with each member having a connection to many other members. It generally expected that online communities are large, sparsely knit, and heterogeneous (Garton et al. 1997). These types of networks are good for obtaining new resources since they have many diverse types of individuals (Rheingold 1993).

Methodology

This study uses the whole network approach to social network analysis by studying relations between members of a population. The population boundary for this study is participants in the Back Pain Support Group bulletin board found at WebMD (**http://boards.webmd.com/topic.asp?topic_id=42**). WebMD was chosen because medical information has been reported as a popular reason that people go online, and WebMD has prominence as a top medical information site (Forster 2002). WebMD claims over 16 million visitors a month (Johnson 2002) and the Back Pain Support Group is fairly active with over 20 messages posted each day. The board also has the minimum required number of members (at least 8 to 12, ideally 20 to 50) to sustain interactive CMC (Jones and Rafaeli 1999). In addition, while links between back pain and demographic variables can be conflicting (Croft et al. 2001; Klaber Moffeett et al. 2000), back pain seems to occur across age groups and genders more readily that other popular health topics on WebMD, such as pregnancy, cancer, and children's issues. Back pain is often a chronic condition, which would prompt the membership of an individual in the community for a longer period than an acute or temporary affliction.

Figure 1. Example of WebMD Back Pain Support Group Messages

All relations between members were collected for one year. A strength of this study is the gathering of actual contact in the social network. Often in social network analysis, the relations are reported by the individuals, which is subject to social desirability bias and reliance on memory. In a bulletin board, initial messages are termed *seeds*. Members may post two types of messages: seeds or replies. A seed together with all of its associated replies is termed a *thread* (see Figure 1). All threads started during the calendar year (including all replies) were collected for this study. As expected, 18.2 percent (2,937) of the total 16,112 messages were seeds, and 81.8 percent (13,175) were replies. For each message, the author, date/time, and subject heading were gathered, as well

as the level of the message (seed or reply). Most importantly, the threads were preserved during data collection. That is, not only the message headers were collected, but also who specifically replied to what message and in what order.

Because the participants could reasonably expect their conversation was being recorded and made public, the data collected was not deemed identifiable private information as defined by the Federal Policy for the Protection of Human Subjects (2001). Further, the WebMD privacy policy clearly states that all bulletin board postings are public information and can be collected by third parties. Therefore, informed consent of the participants was not sought by the researchers. Nevertheless, poster names have been replaced by pseudonyms in reporting the data.

Table 1. Top 25 Posters (Member Posting 1 Percent or More of the Total Messages)

Poster[a]	Number of Posts	Percent of Total Posts	Poster[a]	Number of Posts	Percent of Total Posts
Alpha	514	3.2%	November	250	1.6%
Bravo	453	2.8%	Oscar	228	1.4%
Charlie	429	2.7%	Papa	218	1.4%
Delta	349	2.2%	Quebec	207	1.3%
Echo	340	2.1%	Romeo	203	1.3%
Foxtrot	328	2.0%	Sierra	182	1.1%
Golf	303	1.9%	Tango	180	1.1%
Hotel	302	1.9%	Uniform	171	1.1%
India	299	1.9%	Victor	170	1.1%
Juliet	268	1.7%	Whiskey	163	1.0%
Kilo	267	1.7%	Xray	157	1.0%
Lima	257	1.6%	Yankee	154	1.0%
Mike	250	1.6%			
			Totals (top 25):	6,642	41.7%
			Totals (all):	16,112	100.0%

[a]Poster names have been replaced by the NATO phonetic alphabet to protect their privacy.

A total of 1,670 members posted during the year. Empirical studies have shown that a small percentage of members contribute a proportionately large percentage of messages in virtual communities (Jones and Rafaeli 1999). In addition, it has been found that heavy contributors typically are replying to others in the community rather than starting new threads (Smith 2002). This phenomenon occurs here with one member (Alpha) alone posting 3.2 percent of all messages, while the top 10 posters (only 0.6 percent of total members) posting 22.25 percent of all messages (see Tables 1 and 2). Further analysis will examine if the frequent posters are posting mainly replies or seeds.

Proposed Analysis

For RQ1 (membership stability), the time span in the community will be calculated. Descriptive statistics will show if members stay for longer periods of time, or if people drop in and out. The number of posts per day per user will show participation over time (RQ2). Many communities now show the number of posts made by each member as part of the user's profile. We will examine posts per user to see if there may be natural groups, such as high, medium, and low, based on posting frequency (RQ3). To investigate RQ4 (clusters of members) and RQ5 (core group), we propose social network analysis software such as visone (**http://www.visone.de/**) or Agna (**http://www.geocities.com/imbenta/agna/index.htm**) as well as the multidimensional scaling techniques in SPSS. Similar work has been done in producing social analysis network maps of Usenet newsgroups (**http://web.media.mit.edu/~wsack/CM/**) (Sack 2000) and with Microsoft's Netscan tool (**http://netscan.research.microsoft.com/**). The most interesting diagrams will show the users as nodes and the connections between the users as shorter and closer together the more they conversed. The resulting networks can be produced for different time periods given the longitudinal data. For example, 12 social networks can be drawn, one for each month, to discover how the social structure of the community changes over time.

A matrix has been completed for this analysis with the member names both in the rows and columns. The number of times two people have conversed (defined as appearing in the same thread) will appear at the intersection of the individuals' row and column. Table 3 presents an abbreviated portion of the matrix showing the top 10 posters. This matrix will be used as input to the social network analysis to draw the network diagrams. For example, Alpha and India have "talked" with one another 122 times, so they would appear in close proximity in the network diagram, while India and Juliet have conversed only four times, so they would appear farther from each other. Echo and Golf have not conversed at all, so their nodes would not be connected.

Table 2. All Posters by Frequency Category

Poster Category by Number of Posts Made	Number of Posters in this Category	Percentage of Total Posters	Poster Category by Number of Posts Made	Number of Posters in this Category	Percentage of Total Posters
Over 500	1	0.06	10 – 24	111	6.65
400 – 499	2	0.12	5 – 9	153	9.16
300 – 399	5	0.30	4	66	3.95
200 – 299	10	0.06	3	123	7.37
100 – 199	18	1.08	2	254	15.21
50 – 99	30	1.80	1	853	51.08
25 – 49	44	2.63			

Table 3. Frequency of Interaction Between the Top 10 Posters

Posters → ↓	A	B	C	D	E	F	G	H	I	J
Alpha	0	75	59	70	54	73	48	58	122	21
Bravo	75	0	71	56	17	46	60	89	29	58
Charlie	59	71	0	67	38	56	37	161	42	69
Delta	70	56	67	0	14	80	31	67	42	25
Echo	54	17	38	14	0	24	0	21	34	6
Foxtrot	73	46	56	80	24	0	32	51	44	23
Golf	48	60	37	31	0	32	0	38	5	27
Hotel	58	89	161	67	21	51	38	0	39	83
India	122	29	42	42	34	44	5	39	0	4
Juliet	21	58	69	25	6	23	27	83	4	0

Discussion and Implications

Understanding the pattern of participation in a community through social network analysis can reveal how communities form and grow online. Like most Web sites, sponsors need to know about user retention. Because in virtual communities users interact *with each other* (not just the sponsors of the site) and because the interaction between members produces conversation which is the essence of the Web site, it is also especially vital to understand how people begin to participate, how this participation may change over time, and who they participate with. High turnover in the community, known as *churn*, has been raised as an important issue in the success of virtual communities (Hagel and Armstrong 1997). Classifying users by their posting behavior and examining if posting behavior changes over time will allow virtual community sponsors and researchers to know more about the lifecycle of a community member. Clues about how members enter the community and build up participation should lead to better retention of members ultimately fostering more knowledge exchange in the community.

The existence of a core group in the community is important to know and not readily visible by examining the postings. The core group may be responsible for most of the postings and a majority of replies, and could ultimately be the reason for the community's success or failure. Community sponsors should understand if there is a core group within their community and how

this group may change over time. Clusters of people in the community may influence how the knowledge of the community gets developed and may serve to retain members.

The results of this research may, by understanding who talks to whom, lead to ways of visualizing conversations in virtual communities. Like glancing around a party to see who is speaking to each other in groups, community members could see the composition of groups on their bulletin boards. Visual cues could allow members to draw inferences about the activity in the community and shape their activity (Donath 2002; Erickson et al. 2002). Being able to visualize what others are doing is important in group work and a key reason why face to face interaction is preferable in many cases of communication. Researchers are attempting to add visualization components to virtual communities to in order to support social interaction that is online, showing, for example, when people are reading and responding to others and when they are logged into the community (Erickson et al. 2002). The in-depth analysis of this one community can, in future research, be expanded to different kinds of communities to see how member participation may differ depending on community topic.

References

Butler, B. "Membership Size, Communication Activity and Sustainability: A Resource-Based Model of Online Social Structures," *Information Systems Research* (12:4) 2001, pp. 346-362.

Croft, P. R., Lewis, M., Papageorgiou, A. C., Thomas, E., Jayson, M. I., MacFarlane, G. J., and Silman, A. J. "Risk Factors for Neck Pain: A Longitudinal Study in the General Population," *Pain* (93:3) 2001, pp. 317-325.

Donath, J. "A Semantic Approach to Visualizing Online Conversations," *Communications of the ACM* (45:4) 2002, pp. 45-49.

Erickson, T. "Social Interaction on the Net: Virtual Community as Participatory Genre," in *Proceedings of the 30th Annual Hawaii International Conference on System Sciences*, IEEE Computer Society Press, Los Alamitos, CA, 1997, pp. 13-21.

Erickson, T., Halverson, C., Kellogg, W. A., Laff, M., and Wolf, T. "Social Translucence: Designing Social Infrastructures That Make Collective Activity Visible," *Communications of the ACM* (45:4) 2002, pp. 40-44.

"Federal Policy for the Protection of Human Subjects," in *Code of Federal Regulations*, 2001.

Figallo, C. *Hosting Web Communities: Building Relationships, Increasing Customer Loyalty, and Maintaining A Competitive Edge*, John Wiley & Sons, Inc., New York, 1998.

Forster, S. "Consumers Using Wrong Criteria To Evaluate Medical Web Sites," *The Wall Street Journal*, October 29, 2002, p. D3.

Garton, L., Haythornthwaite, C., and Wellman, B. "Studying Online Social Networks," *Journal of Computer-Mediated Communication* (3:1), 1997 (available online at **http://www.ascusc.org/jcmc/vol3/issue1/garton.html**).

Hagel, J., and Armstrong, A. G. *Net Gain: Expanding Markets Through Virtual Communities*, Harvard Business School Press, Boston, 1997.

Hiltz, S. R., and Wellman, B. "Asynchronous Learning Networks as a Virtual Classroom," *Communications of the ACM* (40:9) 1997, pp. 44-49.

Horrigan, J. B., Rainie, L., and Fox, S. *Online Communities: Networks that Nurture Long-Distance Relationships and Local Ties*, Pew Internet and the American Life Project, Washington, D.C., 2001, (available online at **http://www.pewinternet.org/reports/toc.asp?Report=47**).

Johnson, L. A. "WebMD in the Pink at Nexus of Health Providers, Patients and Payers," *Associated Press Newswires*, November 19, 2002.

Jones, Q., and Rafaeli, S. "User Population and User Contributions to Virtual Publics: A Systems Model," in *Proceedings of GROUP'99 International Conference on Supporting Group Work*, Phoenix, AZ, 1999, pp. 239-248.

Klaber Moffeett, J. A., Newbronner, E., Waddell, G., Croucher, K., and Spear, S. "Public Perceptions about Low Back Pain and its Management: A Gap between Expectations and Reality?," *Health Expectations* (3:3) 2000, pp. 161-168.

Korenman, J., and Wyatt, N. "Group Dynamics in an E-Mail Forum," in *Computer-Mediated Communication: Linguistic, Social and Cross-Cultural Perspectives,* S.C. Herring (ed.), John Benjamins, Philadelphia, 1996, pp. 225-242.

Rheingold, H. *The Virtual Community: Homesteading on the Electronic Frontier*, Addison-Wesley Publishing Company, Reading, MA, 1993.

Rice, R. E. "Communication Networking in Computer-Conferencing Systems: A Longitudinal Study of Group Roles and System Structure," in *Communication Yearbook,* M. Burgoon (ed.), Sage, Beverly Hills, CA, 1982, pp. 925-944.

Ridings, C. M., Gefen, D., and Arinze, B. "Some Antecedents and Effects of Trust in Virtual Communities," *The Journal of Strategic Information Systems* (11:3-4) 2002, pp. 271-295.

Sack, W. "Conversation Map: A Content-Based Usenet Newsgroup Browser," in *Proceedings of the International Conference on Intelligent User Interfaces*, New Orleans, LA, 2000, pp. 233-240.

Smith, M. "Tools for Navigating Large Social Cyberspaces," *Communications of the ACM* (45:4) 2002, pp. 51-55.

Wellman, B. "An Electronic Group is Virtually a Social Network," in *Culture of the Internet,* S. Kiesler (ed.), Lawrence Erlbaum Associates, Mahawah, NJ, 1997, pp. 179-205.

Wellman, B. "For a Social Network Analysis of Computer Networks: A Sociological Perspective on Collaborative Work and Virtual Community," in *Proceedings of the 1996 ACM SIGCPR/SIGMIS Conference on Computer Personal Research*, Denver, CO, 1996, pp. 1-11.

Wellman, B., Salaff, J., Dimitrova, D., Garton, L., Gulia, M., and Haythornthwaite, C. "Computer Networks as Social Networks: Collaborative Work, Telework, and Virtual Community," *Annual Review of Sociology* (22) 1996, pp. 213-238.