Empirical Evidence of Information Overload Constraining Chat Channel Community Interactions

Quentin Jones
New Jersey Institute of Technology, USA
qgjones@acm.org

Mihai Moldovan
New Jersey Institute of Technology, USA
mimold@gmail.com

Daphne Raban
University of Haifa, Israel
draban@gsb.haifa.ac.il

Brian Butler
University of Pittsburgh, USA
bbutler@katz.pitt.edu

ABSTRACT
Prior work has demonstrated that the impact of individual information-processing limits can be observed in dynamics of mass interaction in asynchronous collaborative systems (Usenet newsgroups and email lists). Here we present the first evidence of such impacts on synchronous social interaction environments through the analysis of an Internet Relay Chat network. We highlight how shared public discourse in chat channels appears to be limited to 40 posters in any 20 minute interval, even as the number of channel users increases well into the hundreds. We discuss our findings in terms of understanding the relationship between online community space types and the user interaction dynamics they support.

Author Keywords
Information Overload, Chat, IRC, Synchronous Communication, Online Communities. Computer Mediated Communication.

ACM Classification Keywords
H.5.3. Group and Organization Interfaces: Collaborative Computing.

INTRODUCTION
While more and more web applications rely on user participation and interaction we still have a very limited understanding of the relationship between the design of social computing applications and the development of the communities and groups that they support. This gap in our knowledge relates not just to the newer systems but also to social software in general. The problem is most acute in the case of synchronous social software systems because little work has been undertaken to describe typical patterns of user behavior. For example, although millions of people use Internet Relay Chat (IRC) networks on a daily basis for work, play and education [6], and have done so since the early 1990s [22], little is known about how the groups that make up a Chat network, i.e. the user community of chat channels, develop, grow, and decline. Yet without models of community dynamics and baseline data regarding participation and activity levels for both individuals and groups, evaluating efforts to improve large-scale interaction systems is, at best, problematic. Is it possible to sustain a large, very active group? What is "very active"? What is "large"? How long can a group be expected to last? Do online communities and groups experience a smooth growth trajectory, or are they subject to discontinuities and hard limits? Designers and researchers need answers to these types of questions if they are to successfully evaluate and implement infrastructures to support mass online interaction.

In addition to the problems associated with a lack of basic data and models regarding the development and dynamics of online community interaction, exploring the connection between design and use of social software is complicated by the non-deterministic relationship between human actors, technology, and groups themselves. To side step these problems the tendency has been to examine these issues through qualitative studies which treat each online community as a unique socio-cultural phenomenon [19]. However, while individual cases can provide valuable insight into what is possible, they provide limited information about general patterns of interaction that reflect fundamental relationships between individual behavior, social structure, and technology [9].

The project described in this paper, addresses both the need for baseline data and the need for conceptual models linking individual behavior, technology, and community interaction dynamics. We begin with a presentation of the information-processing constraints model and how it can be used to describe the dynamics of user behavior in online social interaction spaces, such as chat channels. This is followed by a set of hypotheses regarding the dynamics of synchronous online community social interaction spaces. Data from a large-scale study of IRC chat channels is presented both to characterize chat channel dynamics and test the hypotheses derived from the information-processing constraints model. We conclude with a discussion of the findings and their implications for designers and researchers interested in understanding the relationships between different online community infrastructures, individual capabilities, and the dynamics of communities which emerge from the interplay of technology characteristics and user behaviors.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
CSCW'08, November 8–12, 2008, San Diego, California, USA.
Copyright 2008 ACM 978-1-60558-007-4/08/11...$5.00.

Petitioners
Exh-1010, p. 001

INFORMATION-PROCESSING CONSTRAINTS MODEL OF ONLINE COMMUNITY INTERACTION DYNAMICS

Technology, in its most general sense, both enables and constrains mass interaction [26], collective action [18], and large scale collaboration. As space for supporting face-to-face interaction, whether it is a convention center, a manufacturing line, a corporate cubicle farm, or a university campus, makes certain types of interaction easier, while simultaneously making other interactions more difficult. For this reason, understanding how to craft spaces that allow for, and even encourage, desired types of interaction is an important part of any effort to support communities, groups, and organizations.

Historically communication technologies have been developed because they allowed for a broader, or at least, different, set of potential interactions. For example, the ancient Middle East clay tablets with script written on them assisted signal transmissions: they carried information and helped reduce the amount of required interpersonal interactions. Use of this technology helped create many great ancient civilizations. At the same time, clay tablets were also a constraining technology. They were not easily reproduced, and required physical presence for message transfer. Clay tablet technology both enabled and constrained certain social outcomes through its particular form of asynchronous communication.

Similarly, despite the technological advances and popular thinking about the limitlessness of cyberspace, there is substantial evidence that each computer mediated communication (CMC) technology also imposes constraints on the interaction that are supported [13]. CMC technologies allow individuals to communicate over greater distances and across organizational boundaries. Technologies, such as e-mail and web forums, provide a variety of sequencing, timing, and storage options that individuals can choose from as they craft their interactions with others. At the same time, CMC technologies impose costs and constraints on users. How an individual presents themselves to others is often determined, at least in part, by the capabilities of the systems being used. How quickly you can respond, if you can respond at all, is also partially dependant on the capabilities of the CMC systems that are available to support interaction.

The information-processing constraints model of online community interaction dynamics [10],[10],and [13] builds on the idea that CMC technologies enable and constrain individual behavior, by arguing that through both direct, and higher order constraints, these technologies affect the viability and boundaries to the interaction dynamics of online communities.

Communities are formed in order to create and share resources such as information, knowledge, and support. Deriving benefit from these resources requires communication among the community members. In the sustainability model and related empirical results based on listserv data, Butler [2] shows that group size and communication activity have both positive and negative effects. Larger groups attract more members than small groups, but they also lose more members. Similarly, more active groups attract and lose more members than the less active listserv groups. However, while the resource-based theory of online communities acknowledges negative effects (i.e. costs) of group size and activity, it does not explicitly specify how CMC technologies are expected to affect community dynamics.

The information-processing constraints (IPC) model of online communities links technology features and community dynamics by developing the idea that communication technologies are enablers of only a limited range of social interaction dynamics. More specifically, beyond the direct constraints that a technology places on social behavior, the IPC model argues that CMC technology features are central to online communities because they magnify and alter individuals' basic information processing capabilities. The IPC model also highlights the higher-order positive-negative impacts the CMC technologies can have on online community dynamics by removing limits on communication and in the process increasing the information-processing demands on community participants.

Hiltz and Turoff [5] were two of the first researchers to describe how individual information overload results from CMC. Building on their work researchers came to think of information overload as broadly the result of either the delivery of too many communications, that results in individuals receiving more communications than they can easily respond to, referred to as 'conversational overload' [26] or the result of 'information entropy' where incoming messages are insufficiently organized by topic or content to be easily recognized as significant or as part of a conversation's history. Extending these notions to collective action, Van Alstyne and Brynjolfsson [25] argued that the memberships of online communities would Balkanize towards highly specialized groupings as the only reasonable way for individuals to cope with the plethora of groups and millions of internet users. Jones [9] and [11] modified and extended this notion to the nature of the interaction dynamics found in the social interaction spaces that support online communities. He reasoned that once communication load in an interaction space becomes too high, individuals utilizing the shared community spaces will have to employ various coping mechanisms. Further, the cumulative impact of individuals adopting such coping strategies will constrain the interaction dynamics supported by various types of online community spaces.

Itis these higher order impacts, involving the interplay of individual information processing capabilities and technology features, that are central to the information-processing constraints model of online interaction

Petitioners
Exh-1010, p. 002

dynamics. Perhaps the most dramatic implication of this model is the prediction that general human information processing capabilities, combined with the features of CMC technologies will place limits on the range of sustainable interaction structures in online communities. As the number of people in a community increases, the model predicts that the upper limit to individual communication activity (i.e. the user communication density) will decrease due to the increased processing effort required to parse the mix of discourse content, structure and user participation. This prediction is reflected in the dashed upper line in Figure 1.

Figure 1 illustrates how the large-scale mapping of massing interaction dynamics can enable the modeling of the relationship between CMC technologies and communication load [13]. The figure plots an expected relationship between user population, user communication density and technology type. The hypothetical synchronous virtual publics (e.g. IRC chat channels) are located relatively close to the vertical-axis because they require user co-presence for message exchange, which occurs fairly quickly and limits group size. On the other hand, the hypothetical plots of asynchronous virtual publics (e.g. Usenet Newsgroups) have larger user populations because users can take time to digest and respond to messages (which also means interactive message exchange will happen at a slower pace). Of course synchronous CMC systems can be complemented by historical records of interactions that will also impact on users' ability to effectively process interactive discourse [3].

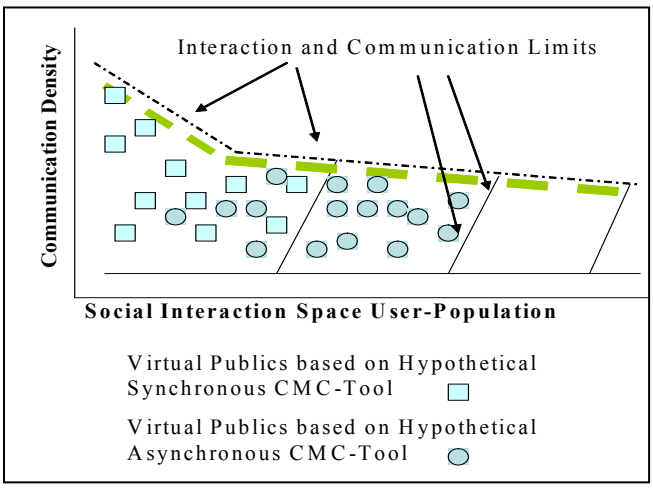

Figure 1. IPC Model of Online Community Viability

The approach / model outlined in Figure 1 does not assume that the technology per se will determine use. Further, it does not suggest anything about the content of the discourse within the boundaries imposed by technology, nor does it prescribe who will use one interaction space or another. Hence, the scatted plots below the predicted upper boundaries to interaction.

Similarly, it is expected that the number of active participants that a community can support will also be subject to a ceiling. Thus, while the number of active participants may initially increase as the overall size of the community increases, it is predicted that communities dependent on synchronous interactions will be subject to a cap, beyond which adding additional active participations is extremely difficult, if not impossible. This prediction of the model is reflected in the vertical lines serve as boundaries on the right side of Figure 1 and the leveling off of the curve in Figure 2 (as discussed in [10] and [12]).

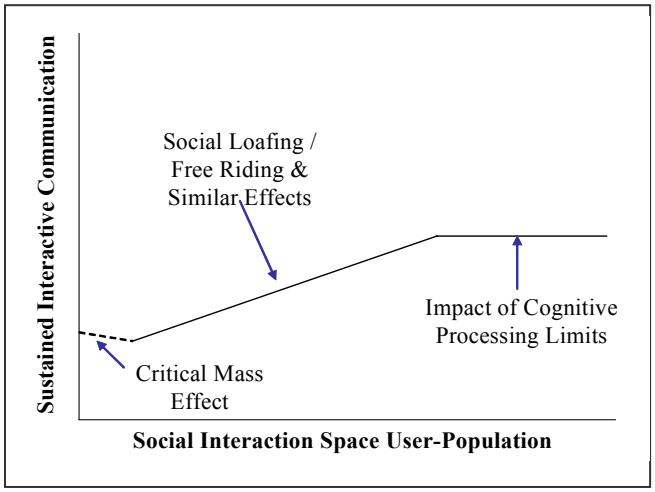

Figure 2. Limits on Growth of Active Participation

At one level the information-processing constraints model, is technology independent. The patterns described above regarding bounds on community interaction are expected to be present no matter what technology is involved. Consistent with this, the model has been successfully used to explore several types of collaborative systems (Usenet and email lists), highlighting the impact of individual information-processing limits on the dynamics of online community interactions [11]. However, while it is expected that the general pattern of limits on community interaction dynamics and community viability will remain, the IPC model also predicts that different technological features will be associated with different boundary parameters (placement and slope of bounding community interaction structures). Hence, while for generalizability purposes it is important to test whether the overall patterns are present when other technologies are used (e.g. synchronous vs. asynchronous), it is also valuable, for both academic and practical purposes, to determine if limits on community viability for different technologies are as predicted (as illustrated by the squares and ovals in Figure 1). Towards this end, the remainder of this paper presents the results of a study of community interaction dynamics in a synchronous communication setting.

Petitioners
Exh-1010, p. 003

INTERNET RELAY CHAT

In the early 1980s various synchronous chat systems emerged that enabled communication organized around spatial metaphors (typically Multi-User Domains or MUDs). These systems were followed by in the late 1980s by chat-channel systems that support conversation between users through the use of a common channel of communication (often known as a "channel", a "chat room", or simply a "room") [24]. The most popular of these was Internet Relay Chat (IRC). IRC supports community communication through channels where a participant can post a line of text (a "turn") to the channel space which is visible to all other participants (As shown in Figure 3 below).

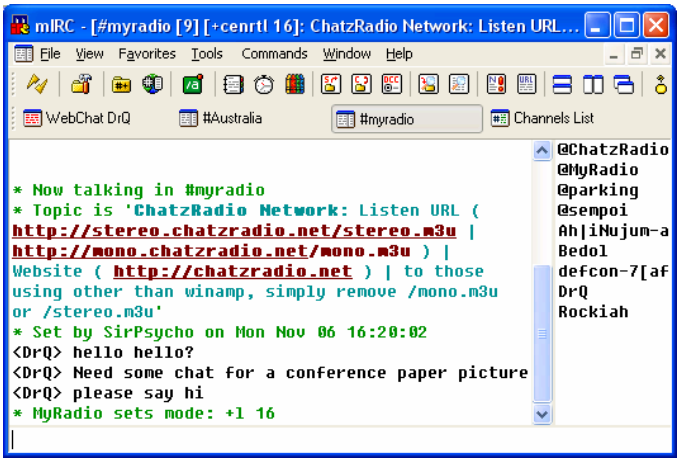

Figure 3. Typical IRC User Interface (MIRC)

The basic chat-channel user interface adopted and popularized by IRC is utilized by many of the other open chat-systems (e.g. Yahoo Chat shown in Figure 4). As a result, gaining an understanding of IRC will also inform our general understanding the thousands of open chat systems found on the Internet.

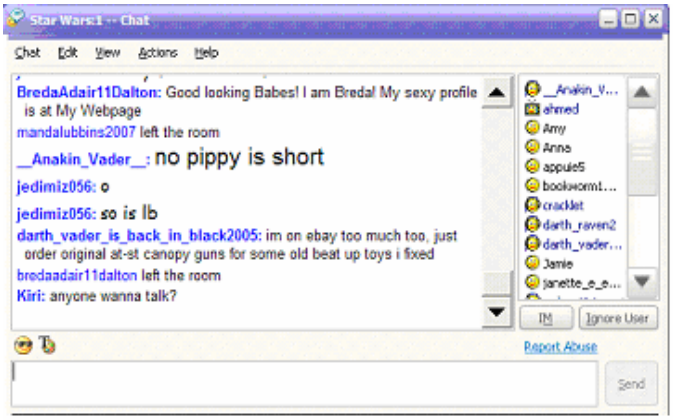

Figure 4. Yahoo Chat User Interface

HYPOTHESIZED LIMITS TO SYNCHRONOUS-COMMUNITY INTERACTION-DYNAMICS

The main research question driving this study is: what is the individual information-processing limit that constrains the community interaction dynamics seen in chat channels? As described above, the core predictions from the information-processing constraints model with regard to these limits relates to the bounds on sustainable communication (or message) density and a cap on the number of active posters that can be involved. The fundamental nature of synchronous CMC with its demand for co-presence for communication leads to the following general hypotheses that will be tested:

H1: As the size of the active chat channel user community grows the maximum average number of chat channel messages posted per individual (message density) will decline to minimal asymptotic level.

H2: The number of active participants of a chat channel (posters) will grow to an observable asymptotic level. Beyond that size, the number of posters co-present in an IRC channel will remain constant

In both cases, the hypotheses describe the presence of a point in the growth of a channel beyond which communication dynamics will change. However, while the IPC model predicts this general pattern it is not possible to hypothesize the specific location of the change point. This will be determined empirically and reported in the analysis.

METHOD

To enable the discovery of constraints to chat channel interaction dynamics we carried out a large-scale field study of the AustNet IRC network. The logic being that by observing a suitably large number of representative channels with varying degrees of activity, we will be able to identify patterns of participation as well as limits thereof. Combining data on activity with data on number of participants will allow us to delineate the community size boundaries.

Data was collected for one year, from February 1 2005 until January 31 2006. During which time we collected 176,966,975 messages, posted to 7,365 active chat channels, by 489,562 unique nicknames. Two methods were used to capture data: 1) IRC bots (software agents) were used that continuously monitored all the channel spaces and collected data at specific time intervals; and 2) through open source TCP traffic monitoring software and associated custom written data parsing system. Both these methods were enabled by our hosting and administering a server on the AustNet IRC network. Users of the network were informed about the research when logging onto the network but the data collection bots and traffic monitoring tools were not visible to network users and did not result in any detectable change in user behavior. Using our administration privileges and customized software we are confident that we were able to reliably distinguish IRC bots

Petitioners
Exh-1010, p. 004

and human users. The data reported in this paper pertains only to human users of the AustNet IRC network.

From this large data corpus set we created a dataset for this study consisting of all observable user actions in 90 representative channels for the month of August 2005 (the middle of the data collection period). The 90 channels were selected by stratified random sampling based on their size (the number of users that join them) and their intensity (how often they were visited during a particular time interval). During this month, there were a total of 2,186 channels with public user postings. Out of these, 1,124 had more than 2 posters, so we selected our sample from this set of channels, eliminating the channels that were not visited by a group of posters (a group being defined as 3 or more users). For each channel in our set we computed the total number of users that visited it during the interval February 1 – August 31 and also the total number of days it was visited by users. Then, a frequencies analysis allowed us to divide the channels into nine categories (three sizes by three intensity levels). We randomly selected 10 channels from each category, resulting in a stratified random sample of 90 channels.

The frequency analysis resulted in the following delimiters for our sampling categories. Small channels were used during the month by between 4 and 98 users; medium channels were used by between 99 and 444 users; and large channels were used by between 445 and 175,141 users. Low intensity channels were visited for less than 45 days from February 1 to August 31; medium intensity channels were visited between 45 and 90 days; and high intensity channels were visited more than 90 visited days.

The raw dataset collected over one month for the 90 channels consisted of 200,880 observations (3 observations per hour x 24 hours x 31 days x 90 channels). The 20 minute observation period was chosen as a reasonable window into short term IRC discourse as 1) the average time per IRC user session, had a modal value of 10 minutes; 2) 50% of users spent 40 minutes or less during an IRC; and 3) a single channel data capture loop for the 2000-6000 channels being monitored took slightly over 15 minutes to complete.

Anonymity, Privacy and Informed Consent

The detailed monitoring and analysis of a large complex social environment such as an IRC network necessitates a discussion of anonymity, privacy and informed consent.

According to the OHRP IRB Guidebook [17], an example of a public place would be "pedestrians on the street", while an example of private space would be "people in their homes". By analogy, we can say that 'open to the internet public' chat channels are public places. As a result Jacobson [8] and others hold that "messages posted to publicly accessible" online environment are "not protected by privacy laws" and the IRB guidebook states that IRB review is not required for "observations of public behavior when the investigator(s) do not participate in the activities being observed". That said we believe that our monitoring of a synchronous communication environment on such a large-scale enables insight into behaviors users would typically consider private and so accordingly we went through a full IRB review.

Users of IRC are generally able to maintain their anonymity even when the system in question is monitored. This is because the two available user identifiers, nickname and IP addresses, are not necessary unique and do not in and of themselves allow for simple conversion/translation to an individual's real world identity. Nicknames can rapidly change and are often interchanged between users and user IP addresses are also change and refer to computer, not people, addresses. To protect user privacy only Institutional Research Board (IRB) approved researchers have been able to access our nickname and IP address data and no publication revealing individual nicknames or IP addresses will be made.

Obtaining explicit informed consent from all IRC users was not possible, and making the effort to do so would have impacted on the ecological validity of our studies. Instead as users logged into the IRC network while we were conducting our research they were presented with information about the research, its aims and data collection methods. This notification along with our decision not to make public any identifiable data (Nicknames and IP addresses and at this stage any conversations) we believe adequately protects the subjects of our study and allows us to conduct ethical chat research [7].

Measures and Analysis Dataset

To describe the dynamics of chat channels and test the proposed hypotheses, it was necessary to construct channel level measures of the number of people involved, the amount of communication activity, and the density of interaction in each channel. We carried out temporal sampling, in which multiple measures were combined to create a single data point for each hour a channel was in use. In other words the raw dataset of 200,880 observations was reduced to 66,960 observations periods (24 hours x 31 days x 90 channels), this was necessary to enable our data analysis. In each case, maximum values observed in our 20 minute were used for this procedure because it best represents the system potential and is the recommended method for analysis of information overload [13]. We decided not to use the first 20 minute period or a mean of the periods as our aim was to use the highest observed level of activity in any given hour, as this would provide the greatest opportunity for the data to falsify our theory. This approach can thus be viewed as a conservative way of assessing the validity of the IPC theory. The following measures were used to characterize the number people involved with each channel:

Petitioners
Exh-1010, p. 005

- Observed Users (OU_max): The maximum number of people who logged into a channel and used it for either private or public conversation during an hour of activity. The measure of the number of observed users for each channel was created by recording the number of users in each 20-minute interval. Then we aggregated the data on an hourly basis by taking the highest of the three relevant counts. The maximum value was selected since it indicates the system potential better than the mean or other measures.
- Observed Posters (OP_max): The maximum number of users who post messages in public to the entire group of channel users during one hour of activity. As with observed users, the measure of the number of observed posters for each channel was created by determining the number of posters present in each twenty minute period and selecting the highest count for each hour.

To describe the interaction activity within a channel the following measure was constructed:

- Observed Messages (OM_max): The maximum number of public postings to an IRC chat channel during an hour of activity. This does not include any messages sent privately between channel participants. As with the previously described measures of observed users and posters, the measure of observed messages was constructed by taking the maximum number of messages observed in the 20-minute sampling frames as an upper bound on the hourly activity.

To characterize the user communication density in an IRC channel the following measures were created:

- User Communication Density (OMperOU_max): User communication density was measured with a ratio of observed messages (OM_max) to observed users (OU_max). This is a measure of message density that is an upper bound on the number of messages per person logged into the each channel.
- Poster Communication Density (OMperOP_max): Poster communication denisity was measured with a ratio of observed messages (OM_max) to observed posters (OP_max). This is a measure of message density that is an upper bound on the number of messages per poster actively participating in each channel.

RESULTS

This study used channel level measures to test predictions for the information-processing constraints model of online communities. It also provides an initial characterization of interaction dynamics within these online community spaces. However, while the analysis is at the channel level, general information about the overall system (the AustNet IRC network) and the individual users help provide context for interpreting the analysis. We will begin with this descriptive background data, and then show the analysis of information overload and channel interaction limits.

IRC system size and users activity

Table 1 shows the average monthly values and total yearly values of a number of characteristics of the AustNet network. It provides a sense of the scale of the interaction environment.

	Monthly Average	Yearly Total
No. of Channels	12,382	61,337
No. of Channels w/users	7,436	31,167
No. of Channels w/posters	2,207	7,181
Total No. of Users	148,564	1,115,147
Total No. of Posters	66,369	489,562
Total No. of Messages	8,621,641	103,459,691

Table 1. Summary Statistics for the AusNet IRC network

Examination of the channel count data (rows 1-3 in Table 1.) highlights how dynamic the environment is with over 7000 channels being used on average each month, with over 2000 of these containing open public discourses.

	Mean	Sample size
Number of channels visited by a poster during a session	2.33	13,043
Number of channels in which a poster was publicly active during a session	1.2	10,648
Time per session spent by a poster on IRC (in minutes)	263.8	57,098
Time until the first posted public message (in minutes)	21.01	33,019
Number of US EST calendar days a poster connected to the network	4.36	64,063
Number of US EST calendar days a poster was publicly active	2.94	64,063

Table 2. Descriptive Statistics of AusNet IRC Network Users

Table 2 provides information on typical users' behavior on the AusNet IRC network system in terms of the number of channels they visit, post public messages in, and the time they spend on the system.

Information Overload

Exploring the raw data it was evident that the values of message density could be divided into several intervals and averaged. This provides an overview of the relationship between channel size and message density. Table 3 reports the mean message density for different numbers of users. It suggests that the rate of decline in message density changes when the number of users reaches about 30, and that 220 is limit on the number of users in an open IRC channel.

Petitioners
Exh-1010, p. 006

Number of Users Per Channel (Hourly)	Mean Message Density (hourly)
2-9	52.99
10-19	21.67
20-29	12.63
30-39	4.51
40-49	4.77
50-59	3.72
60-100	2.69
110-160	2.91
170-220	1.30
220<	0.64

Table 3. Mean Message Density for Different Size Channels

A similar pattern is reflected in Figure 5, the plot of users' hourly communication density (OMperOU_max) versus channel size (i.e. the number of observed users).

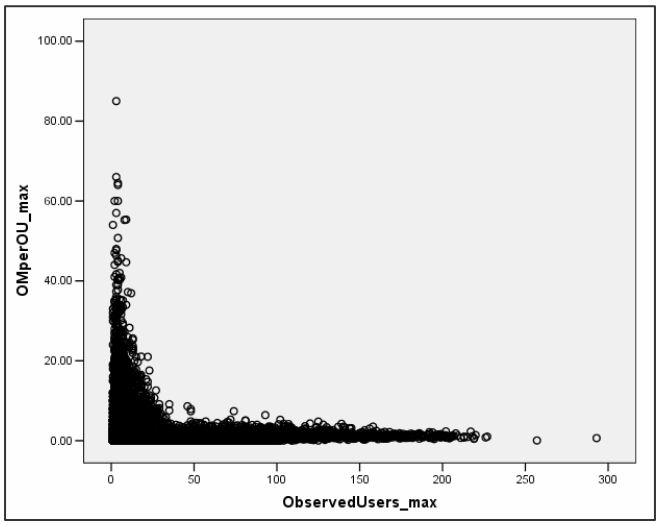

Figure 5. Users' Public Message Density Vs. Number of Users.

This pattern of declining user communication density, to a point, followed by a constant (or much reduced slope) reflects the pattern of community viability predicted by information processing constraints model (Figure 1 and Hypothesis 1).

Figure 6 provides a higher resolution of a similar situation depicting the relationship between hourly message density and the number of observed posters. The pattern in Figure 6 is also consistent with the prediction of the IPC model of online interaction (as reflected in Figure 1).

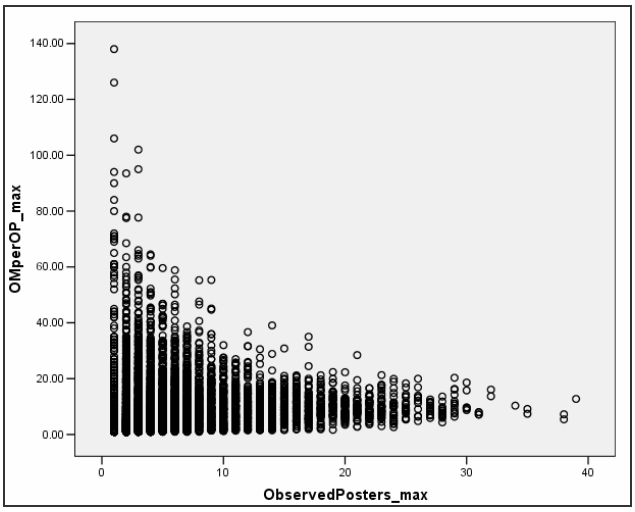

Figure 6. Message Density Vs. Number of Posters.

Again, exploring the data regarding number of posters in each channel, the values of message density were divided into intervals and averaged. Table 4 reports the mean message density for ranges of posters and shows that the limit of active public posters in the system is 39.

Range of Posters	Mean Message Density
2-9	65.97
10-19	29.59
20-29	19.33
30-39	11.73
39<	0

Table 4. Mean Message Density for Ranges of Posters.

Plotting the maximum number of messages versus the maximum number of posters (Figure 7) provided us with the observation that communication activity increases up to a certain point and then appears to reach and asymptotic level as suggested by the IPC model and Figure 2.

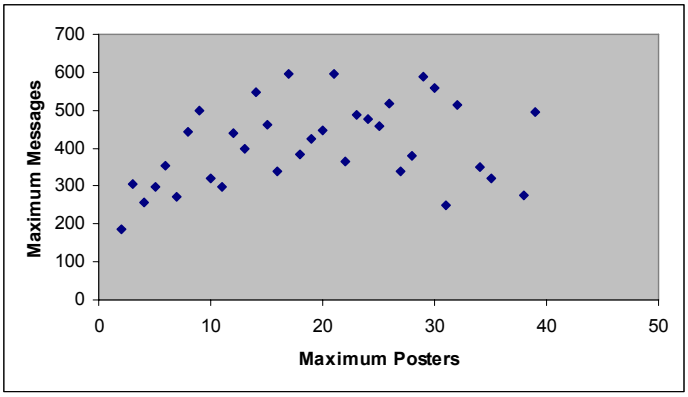

Figure 7. Maximum number of Public Messages Posted Vs. the Maximum Number of Posters.

Petitioners
Exh-1010, p. 007

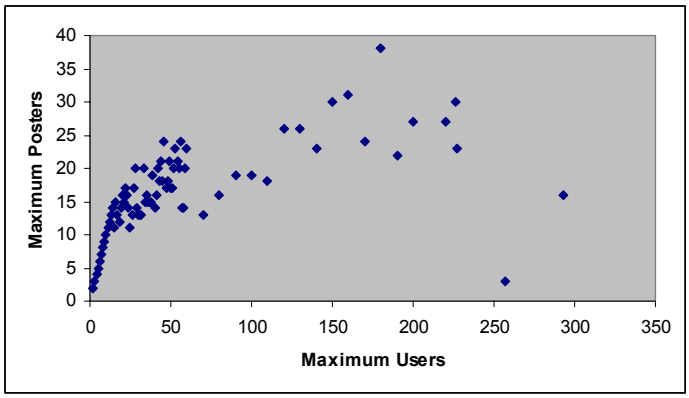

Figure 8. Number of Posters Vs. Number of Users (per channel per hour).

Examining the relationship between the number of observed posters and the number of observed users (Figure 8) provides a basis for assessing the prediction regarding a channel's capacity for active participants (shared/group public channel discourse). Figure 8 shows a linear increase in the maximum number of posters up to a fairly small, number of users. Then the rate of increase diminishes and levels-off between 25 and 30 posters for the range of 150-220 users. Beyond that point (220) there are very few cases of successful channels.

DISCUSSION AND CONCLUSIONS

This paper addresses a fundamental question pertaining to the design of social software and the online interaction spaces they support, namely, what is the relationship between a synchronous CMC technology and user interaction dynamics. Using IRC network data we show that while the system itself does not impose direct restrictions on number of users, posters, or messages, bounds on all three dimensions emerge in ways that are consistent with the information-processing constraints model of online community interaction.

High user message density is possible when the number of users and posters is small as seen in Tables 3 and 4. As the number of participants, users or posters, increases, the user message density declines until community limits are reached. Figures 5 and 7 suggest that, for the studied IRC network, which is an open, synchronous CMC system, the limit of the user community size is less than 300 concurrent users in one chat channel, while the limit of the channel community-space is less than 40 posters (Figures 5 and 7). Figure 8 informs us that the upper limit to the message volume is reached even before the upper limit of posters. A maximum of about 600 public messages per chat channel per 20-minute interval is possible. Roughly, this means that when the poster population is about 30, posters cannot absorb more than 20 messages per poster within 20 minutes or 30 messages per minute for the entire channel group.

As predicted by the information-processing model, the bounds on the rate of posting identified to IRC public channel communication are much lower than those found in asynchronous CMC systems (e.g. the Usenet - see appendix of [13] for details). In Butler's work on email lists we find a mean community size of 168 meaning that the maximum size is much higher. The large pool of potential participants on the AusNet IRC network for chat channel discourse suggests that necessity for co-presence cannot alone explain our observations. Rather, that information processing constraints resulting from information overload are significant factor in shaping the interaction structures present in this otherwise, relatively unconstrained communication environment.

What might be the factors underlying information overload? Why do we see these clear community boundaries? We know there are practical limit to reading, absorbing, and reacting to messages in a given time frame and while we do not have an exact measure of the average communication-processing load associated with various forms of chat channel interaction, we do know that it relates to a number of message system characteristics. Users generally have to make more of an effort to reply coherently to a thread (a chain of messages) than to a single message [13]. Therefore higher message interactivity correlates with higher communication-processing load. Similarly, a dense pattern of messages (high frequency of postings) will require quicker and more sustained processing by group members. Therefore, message density will also co-vary with communication-processing load. Also an increase in interactional-incoherence (e.g. fragmented sentences, disrupted turn adjacency, reversed sequencing, communication lags, agrammatical language, and interactionally disjointed messages) will also increase communication-processing load [4]. It seems logical therefore, that modifications to chat user interfaces that better support the challenges associated with the sequencing of chat discourse (e.g. [24]) would impact on the user population boundaries to chat channel discourse. One would also expect that chat visualization tools could have a similar impact by providing a more persistent view of group discourse [3]. The research outlined not only suggests the likely effect of such modification, but also provides new methods and metrics for their assessment.

Up to a channel size of 14 users the maximum ratio of posters to users is 1 and is seen in Figure 7 as a linear relationship. Beyond 14 users the ratio begins to shrink until about 150 users when it reaches a value of 0.2, which means that only 20% of users are posters. When the ratio is smaller than 0.2 (or 20%) we see that the data becomes less predictable, and the number of observed cases drops significantly. Thus, the data presented confirms that, as predicted by the theoretical model (depicted in Figures 1 and 2) and by our hypotheses, high user message density is possible only at low levels of participation (active or passive). Beyond a certain number of users, about 30 according to Figure 5 and Table 3, user message density remains low and declines until the community loses

Petitioners
Exh-1010, p. 008

viability altogether. This drop in proportional user contribution may result from a form of diffusion of responsibility (or social loafing [14],[15] and [16]): when many people are in a chat channel, specific individuals may feel less inclined to provide responses in the public channel because they know someone else is likely to respond [1][20]. In small groups people may feel more responsible and may also get more credit and appreciation for their participation. All of these factors contribute to the emergent interaction structures resulting from the complex interplay over time between individual's cognitive limitations, technology features and community discourse. As a result, understanding the collective impact of users' information overload coping strategies on interaction dynamics should also inform longitudinal models of chat channel communities [23].

Perhaps the most important aspect of this work is to reengage the CSCW community with legacy chat systems. While IRC is an old technology, it is still used by millions of people around the globe on a daily basis [6] and therefore understanding current user practice is highly relevant to the CSCW community. In addition, as IRC is the progenitor of many of today's synchronous social software user interfaces and protocols, so gaining an understanding of the impact on user behavior of chat channel architecture can inform us more generally about social software design.

ACKNOWLEDGMENTS

This research is partially supported by the National Science Foundation Grant NSF IIS- NSF IIS-HCC- 0749389, CNS 0454081 and DST 0534520. The opinions expressed are those of the authors and may not reflect those of the NSF.

REFERENCES

1. Barron, G. and Yechiam, E. (2002). "Private E-mail requests and the diffusion of responsibility." *Computers in Human Behavior*, *18(5)*, 507-520.
2. Butler, B. S. (2001). "Membership Size, Communication Activity, and Sustainability: A Resource-Based Model of Online Social Structures." *Information Systems Research* 12(4): 346-362.
3. Gergle, D., Millen, D., Kraut, R.E., and Fussell, S.R. (2004). Persistence Matters: Making the Most of Chat in Tightly-Coupled Work. In *Proceedings of the ACM Conference on Human Factors in Computing Systems CHI2004*, pp. 431-438. New York: ACM Press.
4. Herring, S. (1999). "Interactional Coherence in CMC." *Journal of Computer Mediated Communication* 4(4).
5. Hiltz, S. R. and M. Turoff (1985). "Structuring computer-mediated communication systems to avoid information overload." *Communications of the ACM* 28(7): 680-689.
6. Hinner, K. (2000). "Statistics of major IRC networks: methods and summary of user count." *Journal of Media Culture 3(4)*.
7. Hudson, J. and Bruckman, A. (2004). ""Go Away": Participant Objections to Being Studied and the Ethics of Chatroom Research." *The Information Society* **20**: 127-139.
8. Jacobson, D. (1999). "Doing Research in Cyberspace." *Field Methods* **11**(2).
9. Jones, Q. (1997). "Virtual-Communities, Virtual-Settlements & Cyber-Archaeology: a theoretical outline." *Journal of Computer Mediated Communication*. Vol 3., Issue 3., Dec.
10. Jones, Q., (2003). "Appling Cyber-Archaeology." *Proceedings of the Eighth European Conference on Computer Supported Cooperative Work*, Helsinki, Finland. Kluwer Academic Publishers.
11. Jones, Q., and Rafaeli, S. (2000). "Time to split, virtually: 'discourse architecture' and "community building" as means to creating vibrant virtual publics." *Electronic Markets: The International Journal of Electronic Commerce and Business Media* 10(4) 214–223.
12. Jones, Q., and Rafaeli, S. (1999). "User Population and User Contributions to Virtual Publics: A Systems Model." *Proceedings of the ACM International Conference on Supporting Group Work* (Group99). Phoenix, Arizona, USA, ACM Press.
13. Jones, Q., G. Ravid, and S. Rafaeli (2004). "Information Overload and the Message Dynamics of Online Interaction Spaces: A Theoretical Model and Empirical Exploration." *Information Systems Research* 15(2): 194-210.
14. Latane, B. (1981). "The psychology of social impact." *American Psychologist,* 36,4, 357-360.
15. Latane, B., Nowak, A., et al. (1994). "Measuring emergent social phenomena: Dynamism, polarization, and clustering as order parameters of social systems." *Behavioral Science,* 39, 1-24.
16. Latane, B., Wiliams, K. et al. (1979). "Many hands make light work: The causes and consequences of social loafing." *Journal of Personality and Social Psychology* 37,6, 822-832.
17. Office for Protection From Research Risk (1993). "Protecting Human Research Subjects: Institutional Review Board Guidebook." *US Government Printing Office*.
18. Oliver, P., Marwell, G., and Teixeira, R. (1985). "A Theory of the Critical Mass. I. Interdependence, Group Heterogeneity and the Production of Collective Action." *The American Journal of Sociology* Vol. 91 No. 3, pp 522-556.
19. Pacagnella, L. (1997). "Getting the Seats of Your Pants Dirty: Strategies for Ethnographic Research on Virtual Communities." *Journal of Computer Mediated Communication* Vol. 3 No. 1.

Petitioners Exh-1010, p. 009

20. Raban, D.R. and Rafaeli, S. (2007). "Investigating ownership and the willingness to share information online." *Computers in Human Behavior* 23.

21. Rafaeli, S., Raban, D. R., & Kalman, Y. (2005). "Social Cognition Online." In Y. Amichai-Hamburger (Ed.), *The social net: The social psychology of the internet*. Oxford, England: Oxford University Press.

22. Reid, E. (1991). "Electropolis: Communication and Community on Internet Relay Chat." *Department of History Honours Thesis*, University of Melbourne, Melbourne.

23. Schoberth, T., Preece, J. and Heinzl, A. (2003) "Online Communities Longitudinal Analysis of Communication Activities," *Proceedings of the 36th Hawaii International Conference on System Sciences*, IEEE Press, Big Island, Hawaii.

24. Smith, M., Cadiz, J. and Burkhalter, B. (2000). "Conversation Trees and Threaded Chats". *Proceedings of ACM's Conference on. Computer Supported Cooperative Work* (CSCW 2000), Philadelphia PA, USA. ACM Press.

25. Van Alstyne, M.., and Brynjolfsson, E. (1996). "Electronic communities: Global village or cyberbalkans?" *Online, March 1999: www.mit.edu/people/marshall/papers/CyberBalkans.pdf*

26. Whittaker, S., Terveen, L., Hill, W., and Cherny, L. (1998). "The dynamics of mass interaction." *Proceedings of the ACM's Conference on. Computer Supported Cooperative Work.* (CSCW 1998). Seattle, WA, USA. ACM Press.

Petitioners
Exh-1010, p. 010

View publication stats