When is a Group not a Group:
An Empirical Examination of Metaphors for Online Social Structure

Brian S. Butler

August 19, 2003


© Brian S. Butler, 1999
When is a Group not a Group: 
An Empirical Examination of Metaphors for Online Social Structure

Abstract

As extensive computer-mediated communication infrastructures have emerged, both within organizations and in the public sphere, researcher and practitioner interest in networked social structures has increased. One of most common online social structures is the asynchronous electronic collective, in which text-based computer mediated communication systems enable members to broadcast messages to a targeted audience. There are many metaphors that have been applied to these structures, including community, group, forum, and conference. While on the surface these metaphors may seem to be interchangeable, each metaphor is associated with a different set of assumptions about the features and processes of these collectives. Although they have implicitly been the basis for much discussion of these structures, there has been little empirical research that has explicitly compared the various metaphors for online social structure.

A review of field studies of asynchronous voluntary electronic collectives is presented to characterize the metaphors that have been used to describe these social structures. The representations implied by these metaphors are then compared with data from a random sample of e-mail based Internet listservs. In addition, because of the role these metaphors play in discussions comparing traditional and online social structures, pure online collectives and hybrids that combine networked and traditional communication infrastructures are compared. The results indicate that although it is common in studies of computer-mediated communication, the metaphor of 'small groups' does not accurately represent the membership and communication features observed in online social collectives. Furthermore, there is evidence that the characterization of these structures as small groups has biased the existing set of empirical studies. The empirical results suggest that voluntary associations are a more appropriate metaphor, providing a more accurate description and hence better foundation for theorizing about social structures in networked environments.
Networked environments are an increasingly common part of everyday life. Many business, education, and government organizations have invested heavily in the creation of internal communication infrastructures. Similarly, one of the fastest growing segments of the telecommunication industry revolves around the developing public data network known as the Internet. Various systems have been developed within these infrastructures to support social activity. Technologies such as electronic mail and the World Wide Web (WWW) support social activity by allowing members to send messages. Video conferencing and text-based conferencing systems enable individuals who are geographically distant to interact. Whether at work or at home it is more and more likely that people are part of a networked communication system.

Since the early 1980’s, when the earliest computer mediated communication systems were created, researchers have been intrigued by the potential of networked technologies to support, and perhaps change, the way people interact. From this interest has developed an extensive body of research focused on how individuals behave in on-line social environments. In an effort to guide the design of new technologies, much of this work has addressed questions about how social behavior in networked environments might differ from that observed in more traditional face-to-face contexts (e.g. McGuire, Kiesler, and Siegel, 1987; DeSanctis and Gallupe, 1987). In most cases studies have considered how individual behavior in traditional and on-line social contexts compare, and on that basis attempted to infer how the traditional and on-line social structures will differ (Sproull and Kiesler, 1990). However, while the studies of individual behavior in on-line social settings have developed a solid empirical foundation, discussions about the impact of these new technologies on social structure (e.g. Daft and Lewin, 1993) remain based primarily on anecdotes, conjecture, and limited case studies.
Although there are an increasing number of studies that focus on describing examples of on-line social structure, overall this literature provides a weak foundation for theorizing about on-line social structure because it focuses on demonstrating that certain behaviors are possible in networked environments. As computer-mediated technologies developed, the theoretical position that text-based communication media were inherently unsuited for supporting complex social interaction was advanced (Daft and Lengel, 1986). This theory, known as media richness theory, led to the early conclusion that text-based networked environments would be unable to support many types of communication activity. A major thrust of computer-mediated communication research has been to examine the claims of this theory. Whether implicitly or explicitly, past studies of on-line social structures have generally focused on refuting media richness models by documenting the capability of networked environments to support a wide variety of social behaviors (Table 1).

<table>
<thead>
<tr>
<th>Study</th>
<th>Technology</th>
<th>Member Population</th>
<th>Duration</th>
<th>Number of Groups</th>
<th>Primary Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baym, 1993</td>
<td>USENET</td>
<td>Soap opera fans</td>
<td>1 month</td>
<td>1</td>
<td>Participant Observation</td>
</tr>
<tr>
<td>Bikson and Eveland (1990)</td>
<td>E-mail</td>
<td>Corporate employees and retirees</td>
<td>1 year</td>
<td></td>
<td>Survey, Archival</td>
</tr>
<tr>
<td>Collins and Berge (1997)</td>
<td>E-mail Lists (Internet)</td>
<td>Varied</td>
<td>-</td>
<td>8</td>
<td>Survey</td>
</tr>
<tr>
<td>Faraj and Sproull (1994)</td>
<td>USENET</td>
<td>Varied</td>
<td>-</td>
<td></td>
<td>Archival</td>
</tr>
<tr>
<td>Finholt and Sproull (1990)</td>
<td>E-mail Lists (Organizational)</td>
<td>Varied</td>
<td>6 weeks</td>
<td>5</td>
<td>Archival</td>
</tr>
<tr>
<td>Freeman (1984)</td>
<td>Specialized</td>
<td>Social Network Researchers</td>
<td>-</td>
<td>1</td>
<td>Sociometric Survey</td>
</tr>
<tr>
<td>Garramone, Harris, and Anderson (1986)</td>
<td>BBS</td>
<td>Political Constituents</td>
<td>-</td>
<td>1</td>
<td>Survey</td>
</tr>
<tr>
<td>Garramone, Harris, and Pizante (1986)</td>
<td>BBS</td>
<td>Political Constituents</td>
<td>-</td>
<td>1</td>
<td>Survey</td>
</tr>
<tr>
<td>Ha (1995)</td>
<td>E-mail Lists (Internet)</td>
<td>Marketing Professionals and Academics</td>
<td>-</td>
<td>4</td>
<td>Survey</td>
</tr>
<tr>
<td>Hagel and Armstrong (1997)</td>
<td>Internet</td>
<td>Consumers</td>
<td>-</td>
<td>Multiple</td>
<td>Anecdotal</td>
</tr>
<tr>
<td>Study (Year)</td>
<td>Platform</td>
<td>Type</td>
<td>Duration</td>
<td>Participants</td>
<td>Methodology</td>
</tr>
<tr>
<td>----------------------------</td>
<td>---------------</td>
<td>---------------------</td>
<td>----------</td>
<td>--------------</td>
<td>----------------------------------------</td>
</tr>
<tr>
<td>Hiltz (1985)</td>
<td>Specialized</td>
<td>Academic Researchers</td>
<td>1 year+</td>
<td>6</td>
<td>Archival and Surveys</td>
</tr>
<tr>
<td>Hof, Browder, Elstrom (1997)</td>
<td>Internet</td>
<td>Varied</td>
<td>-</td>
<td>Multiple</td>
<td>Anecdotal</td>
</tr>
<tr>
<td>Korenman and Wyatt (1996)</td>
<td>E-mail Lists</td>
<td>Women’s Studies</td>
<td>1 year+</td>
<td>1</td>
<td>Archival and Survey</td>
</tr>
<tr>
<td></td>
<td>(Internet)</td>
<td>Academics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lally (1995)</td>
<td>USENET</td>
<td>MBA Students</td>
<td>-</td>
<td>122</td>
<td>Survey</td>
</tr>
<tr>
<td>Meyers (1987)</td>
<td>BBS</td>
<td>Unspecified</td>
<td>2 months</td>
<td>1</td>
<td>Survey, Archival, and Interviews</td>
</tr>
<tr>
<td>Ogan (1993)</td>
<td>E-mail Lists</td>
<td>Turkish Nationals</td>
<td>1 month</td>
<td>1</td>
<td>Archival</td>
</tr>
<tr>
<td></td>
<td>(Internet)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rafaeli and LaRose (1993)</td>
<td>BBS</td>
<td>Varied</td>
<td></td>
<td>126</td>
<td>Survey</td>
</tr>
<tr>
<td>Rafaeli (1986)</td>
<td>BBS</td>
<td>Students</td>
<td>6 weeks</td>
<td>1</td>
<td>Survey, Archival</td>
</tr>
<tr>
<td>Rheingold (1993)</td>
<td>BBS and Internet</td>
<td>Varied</td>
<td>-</td>
<td>Multiple</td>
<td>Anecdotal</td>
</tr>
<tr>
<td>Rice and Love (1987)</td>
<td>CompuServ</td>
<td>Medical Professionals and Students</td>
<td>6 weeks</td>
<td>1</td>
<td>Archival</td>
</tr>
<tr>
<td>Rice (1982)</td>
<td>Specialized</td>
<td>Academic researchers</td>
<td>24 months</td>
<td>10</td>
<td>Archival</td>
</tr>
<tr>
<td>Roberts (1998)</td>
<td>USENET</td>
<td>Varied</td>
<td></td>
<td>30</td>
<td>Survey</td>
</tr>
<tr>
<td>Rojo (1995)</td>
<td>E-mail Lists</td>
<td>Varied</td>
<td>1 year+</td>
<td>11</td>
<td>Archive and Survey</td>
</tr>
<tr>
<td></td>
<td>(Internet)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smith (1997)</td>
<td>USENET</td>
<td>Varied</td>
<td>3 weeks</td>
<td>4000+</td>
<td>Archival</td>
</tr>
<tr>
<td>Sproull and Faraj (1997)</td>
<td>USENET</td>
<td>Varied</td>
<td>?</td>
<td>&lt; 10</td>
<td>Archival</td>
</tr>
<tr>
<td>Sproull and Kiesler (1990)</td>
<td>E-Mail</td>
<td>Business Organization Members</td>
<td>-</td>
<td>Multiple</td>
<td>Archival</td>
</tr>
<tr>
<td>Sudweeks (1995)</td>
<td>E-mail</td>
<td>Communication Researchers</td>
<td>2 years+</td>
<td>1</td>
<td>Archival, Survey and Interview</td>
</tr>
<tr>
<td>Whittaker (1996)</td>
<td>Lotus Notes</td>
<td>Varied</td>
<td>90 days+</td>
<td>20</td>
<td>Archival and Interviews</td>
</tr>
<tr>
<td></td>
<td>(Organizational)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zenhousern and Wong (1997)</td>
<td>E-mail Lists</td>
<td>Varied</td>
<td>Varied</td>
<td>10</td>
<td>Archival</td>
</tr>
<tr>
<td></td>
<td>(Internet)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 1: Example Studies of On-line Social Structures**

However, while prior field studies have addressed questions about the types of behavior that *can* occur in networked social environments, they have had less to say about what *does* happen. Researchers have typically chosen online sites for study based on personal interest in the content (e.g. Baym, 1995; Ha, 1995) or because the structures were expected to exhibit the social phenomena of interest (Finholt and Sproull, 1990). While these studies are useful existence...
proofs for online behavior, it is unlikely that they provide a realistic description of ‘normal’
operation of online social structures. Similarly, anecdotal accounts are likely to be biased, with
casual observers noticing and reporting interesting events, and providing little or no information
about the features of mundane (or failed) structures. Thus while the studies in this area provide
glimpses networked social structures, they are, at best, a questionable foundation for theorizing
about the development and operation of on-line social strutures.

This study adds to this body of research, developing its basis for generalization by
providing a systematic characterization of a random sample of one type of on-line social
structure, e-mail based Internet listservs. In addition, we also contribute to the study of
networked social environments by empirically comparing features of a set of pure online social
structures with those of hybrid structures that combine networked and traditional infrastructures.
Hypotheses about differences in size, membership change, communication volume, interactivity,
and participation distribution are proposed and tested in order to assess the consequences of
different communication infrastructures for the nature of social structures.

Another influence on on-line social research has been the application of the small group
as a dominant metaphor for characterizing on-line social structures. The metaphors used to
characterize social structures are important because each one embodies a set of assumptions
about the features, processes, and impacts of the structures. Each metaphor partially describes a
social structure, and different models focus attention on different aspects of that structure. For
instance, ‘community’ implies a sense of identity that ‘conference’ does not. Thinking about
‘discussion forums’ suggests that there is extensive interaction among the participants, while
‘mass media’ is likely to have distinct producers and audience members. These are just few
examples of how the selection of a metaphor leads to assumptions about the nature and operation of networked social structures.

Labeling a new phenomenon such as online social structures with a familiar name is useful because it allows researchers to effectively communicate and generalize their findings, by presenting a focused result in the context of a larger framework. Many metaphors have been used to characterize the social structures that have arisen in networked environments (Table 2).

|-------------------------------|-------------|-----------------|----------------|-------------|--------------------------|-----------------------------|

Table 2: Metaphors for On-line Social Structures

However, it is also important to critically examine whether the characterization implied by a metaphor is appropriate. Roberts (1998) and Baym (1993) find evidence of community-like
elements in their studies of USENET groups, supporting the anecdotal reports of Rheingold
(1993) and other popular authors. Finholt and Sproull (1990) and Sproull and Keisler (1990)
report behaviors that are similar to those found in small groups. However, the accuracy of the
metaphors that underlie discussions of online social structures remains largely unconsidered.
Rarely are the prototype structures implied by metaphors compared with one another or with
empirical descriptions of online social structures. Consequently, it is often unclear whether the
assumptions embedded in discussions of online social structure are consistent with the features
and operation of naturally occurring networked social structures.

Many studies of online social behavior have adopted the model of the small groups, and
as a result implicitly assumed that small groups provide an appropriate metaphor for online social
structure. Conceptualizing social structure in terms of small groups provides a theoretical
foundation that makes it logistically and methodologically easier to study the behavior of
individuals in on-line social contexts. Small, task oriented groups communicating synchronously
are easier to recreate in the controlled setting of a laboratory than other social structures which
operate over longer time spans (weeks vs. hours), have less precisely defined goals, and sporadic
participation.

However, while research based on the model of small groups has provided valuable
insights into individual behavior in computer-mediated environments (e.g. McGuire, Kiesler, and
Siegel, 1987; DeSanctis and Gallupe, 1987) it remains unclear whether is it the best foundation
for describing the nature of online structures. As applied in most studies of online
communication, the model of small groups assumes that while the perceptions, attitudes, and
behaviors of individuals may change, the nature of a social structure remains essentially fixed.
Small groups are assumed to be set up, operate, and then they end, typically within a short time
span. Small groups are seen as the context for individual behavior, and not as entities which themselves exist in a larger context. If membership composition is considered at all, it is treated as a causal factor, not an emergent outcome to be explained. Questions are asked about the performance of a group – but not its existance. Likewise, the consequences and causes of membership movement, in the form of new member entry and member loss, received little attention (c.f. McGrath and Hollingshead, 1994: Chapter 5). When considering behaviors, attitudes, and perceptions of individuals in on-line environments it is likely that equating social structure and small groups is appropriate. However, as we move from questions about how individuals behave to questions about how the social structures operate it is necessary to reconsider whether the metaphor of small groups is appropriate, or whether some other model might serve as a better foundation for characterizing on-line social structure.

The analysis presented here examines two alternative metaphors for online social structure and asks which one provides a more appropriate foundation for studies of online social structure. The models, small groups and voluntary associations, were chosen as representative of two broad classes of metaphors used in the exploratory studies of online social structure. ‘Small groups' are most commonly thought of as having fixed, limited membership (< 10 people), high levels of interaction, limited duration, and well defined goals or activities. In contrast, 'voluntary associations', which include social clubs, discussion forums, volunteer organizations, professional societies, conferences, and communities, are expected to have larger, more variable membership; highly uneven, and often non-interactive, participation; extended, if not unlimited duration; and informal, often ambiguously defined, objectives. The appropriateness of these metaphors is tested by comparing structural features seen in a sample of e-mail based Internet listservs, such as membership size and variability, communication volume and structure, and the
distribution of participation, with those expected in a prototypical small group and voluntary association.

**Sample Selection and Data Collection Methods**

The on-line social structures considered in this work are unmanaged, e-mail based Internet listservs. These on-line social collectives\(^2\) utilize Internet-based e-mail and a centrally maintained mailing list to enable individuals to broadcast text-messages to other members. E-mail based collectives were chosen as representative of online social structure because they are known to be prevalent in both private (Finholt and Sproull, 1990) and public network infrastructures such as the Internet.

Although there may be an individual who is responsible to maintaining the mailing list (i.e. the listowner), the selected social structures are unmanaged. Listowners take no formal steps to restrict membership or message content. These collectives are expected to be representative of social structures that operate in environments when there is little active intervention. Hence, the results can be seen as providing a baseline against which the impact of management strategies, such as moderation and member screening, might be evaluated.

**Sample Selection**

The public networked environment of the Internet includes e-mail collectives with various topical emphases, attracting members from a wide range of communities and organizations. From the population of approximately 70,000 collectives, an initial sample of 1066 was created. The initial sample was stratified by topic type to ensure that it spanned a reasonable range of topics and member communities. One third of the sampled collectives

---

\(^2\) The term collective is used to refer to the online social structures. The terms 'group' and 'association' are used to refer to the prototypes implied by the small group and voluntary association metaphors.
focused on work-related topics. One third focused on personal topics (hobbies, lifestyles, etc.). The remaining third involved topics that mixed work-related and personal interests (e.g. geographic locations).

The initial sample was subjected to a multiple stage confirmation process (see Appendix A for more details). Actively managed collectives, including moderated listservs and those with formal new member screening, were eliminated. This selection process also verified that each listserv was mechanically functional, able to provide the needed data, and available for inclusion in the study (See Table 3 for a summary of the reasons for elimination from the sample). The result was a set of 284 listservs, which fell to 204 as collectives were eliminated during data collection.

<table>
<thead>
<tr>
<th>Reason</th>
<th>Number Eliminated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Listowner chose not to participate</td>
<td>227</td>
</tr>
<tr>
<td>Inoperable server or group</td>
<td>120</td>
</tr>
<tr>
<td>Inaccessible membership data</td>
<td>143</td>
</tr>
<tr>
<td>Exclusive membership</td>
<td>86</td>
</tr>
<tr>
<td>Course-related groups</td>
<td>73</td>
</tr>
<tr>
<td>Moderated groups</td>
<td>53</td>
</tr>
<tr>
<td>Broadcast groups</td>
<td>51</td>
</tr>
<tr>
<td>Non-English groups</td>
<td>22</td>
</tr>
<tr>
<td>Sensitive topic/groups</td>
<td>21</td>
</tr>
<tr>
<td>Non-standard message/membership formats</td>
<td>13</td>
</tr>
<tr>
<td>Gateways and non-e-mail lists</td>
<td>9</td>
</tr>
<tr>
<td>Unable to contact the listowner</td>
<td>8</td>
</tr>
<tr>
<td>Incomplete addresses</td>
<td>6</td>
</tr>
<tr>
<td>Duplicates</td>
<td>4</td>
</tr>
<tr>
<td>No description available</td>
<td>2</td>
</tr>
</tbody>
</table>

**Table 3: Reasons for Elimination of a Listserv from Initial Sample**

To verify that the final sample spanned the intended range of topics and populations, the degree to which each listserv’s focused on work-related, personal, or academic concerns was assessed. These measures were constructed by asking coders to read a short description of each
listserv and indicate on a scale from 1 (low) to 5 (high) the likelihood that a substantial portion of each collective’s membership participates for work-related, personal, and academic reasons (three measures for each listserv). Inter-rater reliability was found to be acceptable, with Cronbach alphas of 0.79, 0.88, and 0.78 respectively, and although the sample was not evenly distributed among the three categories, the final sample includes a wide range of topics and membership communities.

Within the final sample listservs were classified as either pure or hybrid collectives. Pure online collectives operate completely in the networked environment. In contrast, hybrid collectives use computer-mediated communication technology to supplement traditional communication activities, such as meetings or print communications (Finholt and Sproull, 1990). Multiple judges were used to assess this feature of each collective. Based on short descriptions, coders assessed, on a 1 (low) to 5 (high) scale, the likelihood that each online collective also used traditional, non-networked, communication activities. Inter-rater reliability was found to be acceptable, with a Cronbach alpha of 0.82. The evaluations were averaged to create a single assessment of each collective's infrastructure. The listservs were then classified either as hybrid or pure based on whether their assessment was above or below the median value for the sample.

Data Collection

For a 130 day period, between July 23, 1997 and November 30, 1997, data on communication activity and membership was collected for each listserv. The communication data consisted of all e-mail messages distributed to members. To collect these messages a project account was created and added to each listserv’s membership list. This account then

---

3 The collected data for listservs eliminated during data collection was archived; however it is not included in the analyses presented here.
received a copy of all messages. The sender identification field was encoded and the messages archived\(^4\). The data collection process resulted in an archive of all communication activity that occurred within the selected collectives during the observation period.

Once a day during the data collection period, a message was automatically sent requesting the listervs' membership lists. As the lists were received, individual contact information was encoded and the data stored. This process generated a record of the membership changes that occurred in the sampled collectives during the observation period. The message and membership archives are the basis of the various measures of collective structure and activity used in the following analyses. Each section will describe the relevant measures and how they were constructed from this raw data.

**Membership Size**

Membership size, as indicated by the number of people who are exposed to a collective's communication activity, is one of the most prominent ways that the metaphors of small groups and voluntary associations differ. Groups are thought of as relatively small social structures, with membership of between 2 and 7 individuals (Forsyth, 1990). Studies of both casual and formal groups have found that group size is distributed according to a j-shaped distribution (e.g. truncated exponential or Poisson distribution), with median values of 2 or 3 (Bakeman and Beck, 1974; Burgess, 1984; Coleman and James, 1961; Desportes and Lemaine, 1988; Dunbar, 1993; James, 1953; Tucker and Friedman, 1972). In contrast, studies of community associations (McPherson, 1983a [Mean: 188, Median: 40]) and youth gangs (Thrasher, 1927 [Mean 31,

\(^4\) Identifying information in both the message and membership data was encrypted in order to address concerns about illicit data use.
Median 16]) found that voluntary associations are larger and that their sizes are log-normally distributed.\footnote{The distribution of voluntary association sizes is also similar to the distribution of business firm sizes (Quandt, 1966; Simon, 1957; Simon and Bonini, 1958; Collins, 1973).}

Collective size is measured by counting the number of members on each listserv's e-mail distribution list on the first day of the observation period. This characterizes size in terms of the number of people who are exposed to the collective's communication activity at that time. While this measure may increase the observed size by counting individuals who receive messages but do not read them, it is conceptually equivalent to counting the number of people who attend traditional meetings, a common measure of size in studies of social structure in non-networked environments.

The distribution of listserv sizes is well characterized by a log-normal distribution (Figure 1).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{membership_size_distribution.png}
\caption{Membership Size Distribution}
\end{figure}
Comparison of the observed distribution of membership sizes with the metaphor prototypes indicates that the online collectives are more similar to voluntary associations than small groups. The listservs are significantly larger than the 2 to 7 range that is expected of groups. However, the mean size of 163 and median size of 64 is comparable to the sizes seen for voluntary associations. Also, as seen for voluntary associations (McPherson, 1983a; Thrasher, 1927) the distribution of membership size among online collectives is log-normal (Figure 1b).

In some discussions of online social activity it has been conjectured that asynchronous computer-mediated communication infrastructures are capable of supporting larger structures than traditional social infrastructures (e.g. Rheingold 1993, Finholt and Sproull, 1990). While this is clearly true when comparing the online collectives (with a mean size of 163 members) with traditional small groups (with sizes in the range of 2 to 7), it is less apparent when comparing them with traditional voluntary associations (with sizes in the 100's and 1000's). Nonetheless, the hypothesis that online collective are larger than traditional voluntary associations is supported by a comparison of the median online collective size (64) and the median size observed in McPherson's (1983a) analysis of voluntary associations within a several Midwest communities (40).

Differences between traditional and networked social environments are also expected to lead to size differences between pure and hybrid social structures. Pure online social collectives exist entirely within the networked environment. As a result, it is argued, they are less affected by the logistical problems that inhibit growth in traditional social environments (Rheingold, 1993; Finholt and Spoull, 1990). In contrast, hybrid collectives combine networked and traditional communication structures, and are more likely to be subject to the costs and
constraints faced by traditional groups and associations (c.f. Hare, Blumberg, Davies, and Kent, 1994; p.147; McPherson, 1983a). Thus, if networked social structures are expected to be larger than those operating in traditional environments, then pure online collectives should be larger than hybrid collectives.

![Membership Size in Pure and Hybrid Social Collectives](image)

Figure 2: Membership Size in Pure and Hybrid Social Collectives

Pure online collectives tend to be larger than hybrid structures (Figure 2). The difference in membership size is significant in the predicted direction (Wilcoxon test: p < 0.001), implying that pure network structures will be larger than hybrid collectives. These results also support claims that networked environments will support larger structures than tradition social environments.

6 The medians were compared instead of means because both sets of data are highly skewed and non-normal.
**Membership Change**

The metaphors of small groups and voluntary associations also differ with respect to membership change. Small groups are seen as having fixed, or at least highly stable, membership. A group is characterized in terms of its members. If the membership significantly changes, it is perceived to be a different group. In contrast, voluntary associations routinely experience high levels of member movement. During the lifetime of an association, many people come and go (c.f. McPherson, 1983a). As a result, both the size and composition of a voluntary association can change significantly over time.

Three measures are used to characterize membership change: percentage growth in membership, percentage loss of members, and percentage change in membership. Percentage growth is the number of new members\(^7\) who arrived during the observation period, relative to a collective’s initial size. Percentage loss is the number of members who left the group over the same time period, normalized by the collective’s initial size. Percentage change combines the measures of member growth and loss to describe the net change in size during the observation period.

Membership change is the norm in the sampled online collectives (Figure 3). More than 75% of listservs had new members during the observation period. Over 50% had members leave.

---

\(^7\) This measure also includes individuals who rejoin after an absence. However, since returning individuals are relatively rare (less than 10% of all recorded 'new' members) no special treatment is given to these individuals.
The online social structures were characterized by significant membership flows, which when operating together resulted in a generally positive change in membership size.

Just as it affects size, the composition of a collective's infrastructure is also expected to affect the rate of membership change. Hybrid collectives, because of they are linked with traditional social infrastructures, should be able to recruit members more effectively. References to the listservs in face-to-face meetings, conferences, or print publications, all raise awareness of the online social activity among a targeted population of individuals who are likely to be interested. In contrast, pure online collectives typically must rely on interpersonal word of mouth or untargeted advertising through the WWW. Thus, hybrid collectives are expected to have higher rate of membership growth than pure online social collectives.

---

8 Annual rates were determined by extrapolating the observed change rates to a 365 day year (365 / 130 * observed value).

---

<table>
<thead>
<tr>
<th>% Growth</th>
<th>% Loss</th>
<th>% Change</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Observed Annual</strong></td>
<td><strong>Observed Annual</strong></td>
<td><strong>Observed Annual</strong></td>
</tr>
<tr>
<td>Mean</td>
<td>22.5%</td>
<td>63.2%</td>
</tr>
<tr>
<td>Std Dev</td>
<td>23.8%</td>
<td>66.8%</td>
</tr>
<tr>
<td>Maximum</td>
<td>117.0%</td>
<td>328.5%</td>
</tr>
<tr>
<td>Q3</td>
<td>34.8%</td>
<td>97.7%</td>
</tr>
<tr>
<td>Median</td>
<td>14.8%</td>
<td>41.6%</td>
</tr>
<tr>
<td>Q1</td>
<td>3.7%</td>
<td>10.4%</td>
</tr>
<tr>
<td>Minimum</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Collectives with 0%</td>
<td>33</td>
<td>53</td>
</tr>
</tbody>
</table>
The connection with traditional communication activity also may affect the rate at which members leave hybrid collectives. Traditional social structures require that members make greater investments of time, energy, and attention, than in pure online collectives. The higher costs make it more likely that the individuals will leave traditional or hybrid collectives than pure networked social structures. Thus because of the costs incurred, pure online collectives are expected to have lower membership loss rates than hybrid collectives.

The expected differences between the membership change processes in pure and hybrid collectives were not observed in the listserv data (Table 2).

<table>
<thead>
<tr>
<th></th>
<th>Pure (N = 131)</th>
<th>Hybrid (N = 73)</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Growth</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>21.9%</td>
<td>23.2%</td>
<td>-1.3%</td>
</tr>
<tr>
<td>Median</td>
<td>14.8%</td>
<td>16.7%</td>
<td>-1.9%</td>
</tr>
<tr>
<td>% Loss</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>13.5%</td>
<td>13.7%</td>
<td>-0.2%</td>
</tr>
<tr>
<td>Median</td>
<td>8.1%</td>
<td>8.1%</td>
<td>0.0%</td>
</tr>
<tr>
<td>% Change</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>8.4%</td>
<td>10%</td>
<td>-1.6%</td>
</tr>
<tr>
<td>Median</td>
<td>4%</td>
<td>2%</td>
<td>2.0%</td>
</tr>
</tbody>
</table>

Table 2: Membership Change in Pure and Hybrid Social Collectives

Although the growth rates differ in the predicted direction (i.e. the hybrid collective's growth rate is higher) the difference is not statistically significant (Wilcoxon, p > 0.1). There is also no significant difference between the membership loss rates for the two collective types.

Communication Activity

Communication among members underlies coordination, social support, information sharing, and other social process, such as identity or norm formation, which are essential to the operation of any social structure. Yet the amount and structure of communication implied by the metaphors of small groups and voluntary associations differ. Small groups are seen engaging in
limited sessions involving high levels of interactive communication. Some theorists have
defined small groups as collections of individuals\(^9\) who influence one another through interaction
(for review see Forsyth, 1990: pp. 6-8), highlighting the importance of communication in these
social structures. The expectation with small groups is that they involve members in a limited
session with high levels of communication activity. In contrast, voluntary associations, with their
long lifespans, are expected to involve a lower volume of communication activity, often making
use of structured meetings, informal gatherings, and print media to maintain communication
among the members.

Communication activity volume in online collectives is measured in terms of the average
number of messages per day. In a listserv, each message represents a member taking a 'turn' in a
conversation. Among the sampled collectives there is significant variation in the communication
volume. However, it does not appear that the norm is high levels of activity. One third of the
listservs had no communication activity during the observation period (Figure 4).

---

\(^9\) Forsyth (1990) is an example of researcher who focus on groups as social structures. There is also a body of
research that conceptualizes groups as psychological constructs. These “minimal group” studies are based on the
Among the online social collectives that were active during the observation period, the mean daily communication volume is concentrated at the low end, with a median that is the equivalent of one message every 3.6 days.\textsuperscript{10}

Another difference between the small group and voluntary association metaphors is the expectations regarding the distribution of communication activity. Small group sessions are assumed to be communication oriented, to the point that a collection of people who came together but did not talk to one another would probably not be considered a group (Forsyth, 1990). This assumption leads to the characterization of small groups as having high levels of ongoing communications activity. In contrast, voluntary associations are characterized as having relatively uneven communication flows. For example, the amount of communication activity in a professional organization might be low with 'bursts' of activity occurring around intermittent meetings, conferences, and print publications.

To characterize the distribution of communication activity in the online social collectives, a Gini coefficient was calculated with each day as a category. The Gini coefficient is a value between 0 and 1 (inclusive) which describes the concentration of items in a set of categories. A low value indicates that items (e.g. messages) are spread evenly across the categories (e.g. days). A high value indicates that they are highly concentrated, with a few of the categories (e.g. days)

---

\textsuperscript{10} The distribution of mean daily communication activity among the active online social collectives has the following features:

<table>
<thead>
<tr>
<th>N</th>
<th>Mean:</th>
<th>Std Dev.</th>
<th>Maximum</th>
<th>Median</th>
<th>Minimum</th>
</tr>
</thead>
<tbody>
<tr>
<td>136</td>
<td>1.635</td>
<td>3.543</td>
<td>29.121</td>
<td>0.277</td>
<td>0.007</td>
</tr>
</tbody>
</table>

---
accounting for a large number of the items (e.g. messages). This provides an overall measure of the degree to which communication activity seen during the observation period is evenly (or unevenly) distributed.

In online social collectives, communication activity is not evenly distributed over the observation period (Figure 5).

![Daily Concentration of Communication Activity](image)

<table>
<thead>
<tr>
<th>N:</th>
<th>136</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum:</td>
<td>1</td>
</tr>
<tr>
<td>Mean:</td>
<td>0.751</td>
</tr>
<tr>
<td>Median:</td>
<td>0.833</td>
</tr>
<tr>
<td>Std Dev.:</td>
<td>0.227</td>
</tr>
<tr>
<td>Minimum:</td>
<td>0.205</td>
</tr>
</tbody>
</table>

**Figure 5: Communication Activity Concentration**

The mean value of 0.751 and the left-skewed distribution among the sampled collectives indicates that communication activity in these social structures tends to occur in a few bursts rather than evenly over time. Overall, the online social collectives are best characterized as having low volumes of highly concentrated communication activity.

The expected effect of combining networked and traditional modes of communication on a collective's communication activity is unclear. Some work suggests that communication activity will be greater in hybrid contexts and other results imply that online social activity will greater in pure networked environments. E-mail is often used in organizational settings to coordinate activities and share information in support of other off-line communication activities.
(meetings, project collaborations, etc.). The presence of a relationship supported by face-to-face communication is expected to increase the ability of individuals to use text-based communication media. A significant relationship has been found between who people interact with in traditional settings and who they communicate with via E-mail (Rice, Grant, Schmitz, and Torobin, 1990). For this reason hybrid infrastructures might be expected to see higher levels of communication activity than pure collectives.

On the other hand, hybrid structures operate in a context that provides members with alternative means for interacting as a collective (Finholt and Sproull, 1990). Members of hybrid collectives have multiple communication media to choose from, while the participants in pure online collectives have little choice but to use the networked communication tools. To the degree that communication media are substitutes, the availability of traditional communication opportunities may reduce use of the online communication. In contrast to the above argument, this characterization of online communication implies that hybrid social collectives will have lower volumes of online social activity than pure networked collectives.

Although graphically there is some evidence that activity in pure online collectives may be greater than in hybrid collectives (Figure 6), the difference is not statistically significant (Wilcoxon Test: $p > 0.1$).
Pure and hybrid collectives also differ in terms of the proportion of structures that saw no activity during the observation period (Table 3).

Table 3: Proportion of Online Collectives with No Communication Activity

<table>
<thead>
<tr>
<th></th>
<th>No Activity</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hybrid</td>
<td>29 (40%)</td>
<td>44 (60%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>73</td>
</tr>
<tr>
<td>Pure</td>
<td>39 (30%)</td>
<td>92 (70%)</td>
</tr>
<tr>
<td></td>
<td>68</td>
<td>136</td>
</tr>
<tr>
<td></td>
<td></td>
<td>204</td>
</tr>
</tbody>
</table>

However, a Fisher exact probability test (p = 0.165 > 0.1) indicates that there is no significant relationship between the collective type (hybrid vs. pure) and the proportion of collectives that see no activity. These results suggest that the presence or absence of traditional infrastructure elements does not significantly affect the volume of communication activity in networked social structures.

11 The data point with this maximum value (29.131) was excluded from the figure to allow for more effective comparison of the category distributions.
Group Communication Structure

Small groups and voluntary associations also differ in terms of the structure of communication activity. Small groups are generally seen as being interactive, with members taking turns in an ongoing stream of interrelated conversation (Bonito, 1997; Hollingshead and Bonito, 1998). In these contexts, individual members hear and respond directly to the comments of others. In contrast, communication activity within voluntary associations is expected to be more episodic. Although there are still likely to be themes and topics that are common throughout the stream of communication, because of logistical and temporal constraints, there are significantly fewer explicit responses.

Message activity in online collectives has the potential to have an interactive structure. Discussion threads, formed by a set of messages that share a common subject line, are a common communication structure that is considered to be indicative of interaction in asynchronous online environments (Sproull and Faraj, 1997). The proportion of a collective’s messages which receive no reply (i.e. solo messages) and the average number of messages within a discussion thread, (including solo messages as threads of length 1) are thus two values which provide an indication of the level of interaction (Sproull and Faraj, 1997). These measures characterize the ‘public’ interactivity or the explicit structural interaction present within the group communication. They do not capture interaction that takes place through traditional communication media, personal e-mail outside a collective's communication infrastructure, or members’ perceptions of interactivity (Koreman and Wyatt, 1996). However, as Finholt and Sproull (1990) note, communication features such as these are important to consider because they are highly visible, and hence are likely to play a significant role in individuals’ perception and behaviors.
Solo messages are identified by first categorizing each message as either a new message or a reply to an earlier message. This categorization is performed based on the contents of the message subject line. Subject lines that begin with ‘re:’ are classified as replies\(^\text{12}\). All other messages are categorized as new messages. Discussion threads are identified by removing the ‘re:’ and matching the first 40 characters of each reply subject line with the subjects of previously distributed new messages. Thread length was determined by counting the number of messages within each identified thread. Solo messages are discussion threads that have a length of one. The proportion of solo messages was computed by dividing the number of single message threads by the total number of messages distributed within the collective during the observation period. This value provides a measure of the interactivity of a collective’s communication, with lower proportions of solo messages indicating higher interactivity. Average thread length also serves as a measure of interactivity. An online collective with shorter threads sees relatively less public interaction while longer threads indicate that the communication activity regularly includes explicit interaction.

Overall, 32% (10439 ÷ 32373) of the recorded messages are solo messages. This suggests that, within the sampled collectives, extended public interaction is somewhat unusual. The distribution of thread length also supports this characterization with at least 75% of the observed threads involving only 1 or 2 messages (Figure 7).

\(^{12}\) To test the reliability of this classification rule a sample of 500 messages were classified manually. The error rate of the automatic classification rule (‘re:’ in the subject line) was less than 1%.
Figure 7: Distribution of Discussion Thread Lengths

The general lack of explicit interaction is also reflected in the measures of collective interactivity.

In over half of the sampled collectives, a majority (> 50%) of the communication activity was solo messages (Figure 8).

Figure 8: Distribution of the Proportion of Solo Messages
Also, thread lengths within the online collective was short, with most (more than 75%) of the listservs having mean thread lengths of less than two messages (Figure 9).

Figure 9: Distribution of Mean Discussion Thread Length

The sampled online collectives are not highly interactive, at least in terms of the structure of the group communication. A significant portion of all messages are solo messages (32%) and a majority of the structures have more than half of their communication in the form of solo messages. Overall thread lengths are short (1 or two messages) and most online collectives can be characterized as tending to have short public discussions (<= 2 messages).

In hybrid groups the availability of alternative communication opportunities is likely to affect the structure of communication activity. In many cases, complex interaction can be accomplished more efficiently and effectively in a face-to-face setting. Thus, the availability of face-to-face interaction should reduce the average complexity of online communication, resulting in shorter messages and less explicit interaction in the hybrid collectives. In addition, common activities, experiences, culture, and shared physical spaces all provide mechanisms for more
efficient communication. Shared knowledge of a physical space, for example, allows things to be referenced, and hence discussed, more succinctly. The combined effect of shared context and the availability of alternative communication opportunities suggests that social activity in hybrid collectives should be more compact and less interactive than in pure networked collectives.

A significant difference in the mean message length in the two sub-populations supports the prediction that activity in hybrid collectives will be more compact than in pure online collectives (Wilcoxon test: p = < 0.0001) (Table 4).

<table>
<thead>
<tr>
<th></th>
<th>Pure</th>
<th>Hybrid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Message Length (Number of Words)</td>
<td>403 (N=23,906)</td>
<td>312 (N=8,467)</td>
</tr>
</tbody>
</table>

**Interactivity**

<table>
<thead>
<tr>
<th></th>
<th>Pure</th>
<th>Hybrid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportion of Solo Messages</td>
<td>60% (N = 94)</td>
<td>61% (N = 44)</td>
</tr>
<tr>
<td>Average Thread Length</td>
<td>1.61 (N = 94)</td>
<td>1.50 (N = 44)</td>
</tr>
</tbody>
</table>

**Table 4: Communication Structure in Hybrid and Pure Online Collectives**

However, while there is a small difference in the predicted direction for the average thread lengths in hybrid and pure collectives, it is not statistically significant (1.50 vs. 1.61: Wilcoxon test: p > 0.1). There is also no difference between the hybrid and pure collectives in terms of the proportion of communication activity accounted for by solo messages (61% vs. 60%). Thus while the composition of a collective's infrastructure may affect features of individual messages, there is no evidence that it significantly alters the overall structure of the communication stream with respect to public interactivity.

**Participation Patterns**

Both small groups and voluntary associations are known to exhibit uneven participation distributions. In both small groups and voluntary associations, it is common for a small
percentage of membership to account for a majority of the communication activity (Bales, Strodtbeck, Mills, and Rosenborough, 1951; Warner and Hilander, 1964; Skvoretz, 1988). However, small groups and voluntary associations differ in the degree to which active participation is unequal. Within small group sessions it is not uncommon for the top one or two active participants to account for 50-75% of the communication activity (Bales, Strodtbeck, Mills, and Rosenborough, 1951; Skvoretz, 1988), while the least active members contribute relatively little. However, while there is clearly an unequal distribution of activity, it is generally not the case that a substantial portion of the membership is silent. In contrast, within voluntary associations large segments of the membership may be passive participants, not contributing at all to the communication activity (Warner and Hilander, 1964; Warner, 1965). In these structures, there is typically a pronounced dichotomy between the active core members and the silent periphery (Lyon, 1974).

The participation structure of the online collectives has been characterized in terms of three measures: participation ratio, a Gini coefficient for the distribution of participation among the active participants, and the proportion of activity accounted for by the two most active members. The participation ratio, or proportion of members who contribute at least one message (i.e. active members), captures the degree to which the membership as a whole is actively involved in a collective's on-line social activity. The concentration of communication among the active participants was described with a Gini coefficient calculated for each online collective. The Gini coefficient is a value between 0 and 1 (inclusive) that indicates the degree to which messages are concentrated among the active participants. A low value indicates that communication is equally distributed among the active participants and a high value indicates
that it is concentrated, with a few individuals contributing most of the messages\textsuperscript{13}. The
distribution of communication among the participants was also measured by determining the
proportion of communication activity that is accounted for by the two most active participants.
This measure was constructed because it was expected that the features of the high end of the
participation distribution have been highlighted in prior research on groups and associations, but
the Gini coefficient is known to be insensitive to differences at the extremes (Smolensky, 1994).
For comparison these measures were also calculated for a set of 30 traditional small groups
working on a discussion task (data was originally reported in Skovretz, 1988).

The participation structure of online social collectives tends to be highly concentrated
among a relatively small number of members. The participation ratios are skewed to the right
with most of the collectives having fewer than 20\% of the membership actively participating
(Figure 10).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{participation_ratio_distribution}
\caption{Participation Ratio Distribution\textsuperscript{14}}
\end{figure}

\textsuperscript{13} The range of possible Gini coefficient values is not $[0,1]$. Use of only active members (i.e. whose who sent at least
one message) ensures that the extreme situation (i.e. all messages being sent by one person with the other included
individuals sending no messages), and hence the extreme value of 1, cannot occur.

\textsuperscript{14} Six cases with participation ratios greater than 100\% are not included in this figure. This occurred in several small
collectives that experienced significant membership growth during the observation period.
Likewise, among the active participants the activity was not evenly distributed (Figure 11)

(a) Gini coefficient

(b) Proportion of the activity accounted for by the top two participants

Figure 11: Distribution of Participation Concentration
In comparison to the Skvoretz (1988) small group data, the online social collectives have more concentrated participation patterns. In contrast to the 30 traditional small groups studied by Skvoretz, in which the participation ratios were all higher than 80%, the online collectives had a majority of members who did not actively participate. Furthermore, among the active participants, the small groups had a participation pattern that was more equally distributed than the pattern seen online (Figure 11a). However, while in the online collectives participation was overall more concentrated, at the extreme the difference was reversed with the top participants accounting for a lower proportion of the activity in the online collectives than in the small groups (Figure 10b).

In prior work it has been argued that one of the valuable features of networked environments is their potential to equalize participation (Dubrovsky, Kiesler, and Sethna, 1991). By eliminating the logistical problem of blocking (Gallupe, Dennis, Cooper, Valacich, Bastianutti, and Nunamaker, 1992) and reducing the social cues (Sproull and Kiesler, 1986), these technologies are expected to reduce the factors which can lead to participation inequality (Skvoretz, 1988). However, subsequent research has shown that participation differentials can persist due to status differences (Saunders, Robey, and Vaverek, 1994; Weisband, Schnieder, Connolly, 1995) and differences in individuals’ expectations regarding participation (Rojo, 1995). As a result, the distribution of participation in online social collectives can remain concentrated among a small subset of the members.

Hybrid collectives are more likely to be subject to known status differences among the members, an important aspect of the process by which status is linked to participation (Weisband, Schnieder, and Connolly, 1995). Therefore, it is expected that
participation in hybrid collectives will tend to be more concentrated than in pure online collectives.

The differences observed in these results suggest that while hybrid groups may be more concentrated among the active members, a greater proportion of their members actively participate in online social activity (Table 5).

<table>
<thead>
<tr>
<th>Participation Ratio</th>
<th>Pure Online Collectives (N= 94)</th>
<th>Hybrid Collectives (N = 44)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean 28%</td>
<td>52%</td>
</tr>
<tr>
<td></td>
<td>Median 15%</td>
<td>19%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Concentration among active members (Gini coefficient)</th>
<th>Pure Online Collectives (N= 72)</th>
<th>Hybrid Collectives (N = 36)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean 0.3815</td>
<td>0.3920</td>
<td></td>
</tr>
<tr>
<td>Median 0.3765</td>
<td>0.3957</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Proportion of Messages Sent by Two Most Active Participants</th>
<th>Pure Online Collectives (N = 72)</th>
<th>Hybrid Collectives (N = 36)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean 36%</td>
<td>42%</td>
<td></td>
</tr>
<tr>
<td>Median 31%</td>
<td>37%</td>
<td></td>
</tr>
</tbody>
</table>

Table 5: Participation Concentration in Hybrid and Pure Online Collectives

However, while these results suggest that there are differences between the participation structures in hybrid and pure online collectives, they are not statistically significant (Wilcoxon test: p > 0.1 in all three cases).

Discussion

These results imply that listservs are best described as large, dynamic social structures in which a core of active participants generates relatively low levels of sporadic communication (Table 6).

---

15 The number of listservs in each condition varies due to differences in the number of participants. For example, listservs with no message activity are excluded. Likewise, participation concentration measures (e.g. the Gini coefficients) cannot be calculated for listservs with only one participant.
<table>
<thead>
<tr>
<th>Membership</th>
<th>Small Groups</th>
<th>Voluntary Associations</th>
<th>Online Collectives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>3-10</td>
<td>30+</td>
<td>44</td>
</tr>
<tr>
<td>Growth and Loss</td>
<td>Little or none</td>
<td>Constant</td>
<td>Significant</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Communication Activity</th>
<th>High</th>
<th>Low</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volume</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Distribution</td>
<td>Constant</td>
<td>Sporadic/Bursty</td>
<td>Sporadic/Bursty</td>
</tr>
<tr>
<td>Structure</td>
<td>Interactive/Conversational</td>
<td>Episodic</td>
<td>Episodic</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Participation Structure</th>
<th>Full membership</th>
<th>Dichotomous – Active Core and Passive Periphery</th>
<th>Dichotomous – Active Core and Passive Periphery</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distribution among Active</td>
<td>Concentrated</td>
<td>Concentrated</td>
<td>Concentrated</td>
</tr>
<tr>
<td>Participants</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 6: Comparison of Small Groups, Voluntary Associations, and Online Collectives**

In terms of membership size and change, communication volume and structure, and participation Internet listservs are more like voluntary associations than small groups. These findings highlight a bias in prior studies of online social activity. While the goals of verifying the existence of recognizable social behaviors in networked environments have been well served by focusing on highly active, interactive examples of online social structures, at least for e-mail based Internet structures, these cases do not seem to be representative. For example, interactivity is a common theme in many descriptions of on-line social activity (Rheingold, 1993; Baym, 1995; Hof, Browder, and Elstrom, 1997). Cases often highlight different types of interaction that can occur in e-mail based social contexts (Sproull and Kiesler, 1990; Finholt and Sproull, 1990). However, the results presented here imply that while interactivity can occur in these contexts, its is more the exception than the norm. Another feature that is common to most of the structures described in prior work is a reasonably high volume of communication activity. In some cases this is acknowledged as an explicit selection criteria (Finholt and Sproull, 1990) while in others it is a result of researchers desire to work with clearly visible social phenomena (e.g. Rheingold,
1993; Baym, 1993; Ogan 1993). However, while this strategy is effective for documenting the
types of social processes that can occur in networked environments, the results presented above
imply that prior work may have unintentionally presented a biased description of the social
activity that is likely to occur in a networked social environment.

This bias is significant because of the effect that it has on discussions, both academic and
popular, about online social structures. For example, contrary to discussion of the problems of
developing electronic social collectives which focus on minimizing the consequences of
unwanted communication (e.g. Kollock and Smith, 1996; Kollock, 1997), these results imply that
a major problem facing developers is prompting some appropriate level of communication. That
is, while “free riding” behavior, in which individual contribute unwanted messages, may be a
problem in some cases, it seems that a more common problem is collective silence. From a
practical standpoint, this implies that rather than focusing on controlling contributions,
developers should devote their attention to encouraging participation. More fundamentally, this
suggests that to better ground the discussion of networked social environments additional work
should be done to document the characteristics of online social activity and structures in a variety
of contexts.

Clearly one limitation of this work is that it only considers one type of online social
structure: listservs. While listservs can be considered representative of a large class of on-line
social structures, including WWW conferencing systems, USENET, and other structures based
on asynchronous communication infrastructure, it is possible that other on-line social structures
might be more "group-like". Specifically, on-line social structure which make use of
synchronous communication technology, such as Multi-User Dungeon (MUDs) or Chat rooms,
might be expected to have features (size, communication activity, participation, etc.) more like
small groups and less like voluntary associations. Also, while the comparison of pure and hybrid listservs revealed few differences, it is also possible that online social structures existing within the context of a single organization might also operate differently. Additional population studies would further the study of computer-mediated communication by providing a better understanding the types of social structures which arise within different communication infrastructures.

These results also call attention to the assumptions underlying discussions of "new" forms of organizing. Technology impact claims are made with respect to some baseline. Discussions of technology enabling "new forms" of social structure, implicitly make assumptions about what "old forms" of social structure looked like. If the metaphor of small groups is used then the baseline is likely to be a prototypical traditional small group. This leads to the conclusion that in most cases on-line social structures are indeed a new form of organizing. However, if the baseline of voluntary association then the validity of the novelty claim is less apparent. As the analysis of pure and hybrid online collectives indicated, there is evidence for some differences between social structures that make use of different communication infrastructures. However, overall, the listservs had membership, communication, and participation features that were generally similar than those expected from the prototypical voluntary association. Thus claims about the impact of communication technology on social structure may over state the novelty because they implicitly compare apples (online social collectives) and oranges (traditional small groups), rather than two types of apples (online social collectives and traditional voluntary organizations).

A radical interpretation of these findings would suggest that small groups should not be used as a foundational metaphor for the study of on-line social structure. The conceptual
framework underlying small group research has embedded in it assumptions about size, membership stability, levels of communication activity, interactivity, and participation structures. However, while these assumptions may be adequate for characterizing the context for examining individual social behavior it cannot be assumed that they are appropriate when the structures is itself the object of study. Thus, while studies of online social activity based on the small group paradigm can provide valuable insight into individual behavior in online social context, applying that model in discussions about the operation of naturally-occurring networked social structures must be done critically, if at all.

In contrast, conceptualizing many on-line social structures as associations or organization may be more appropriate than seeing them as meetings or social gatherings (i.e. small group sessions). While there are technologies, such as Chat Rooms and MUDs, which enable synchronous communication session, the most common communication infrastructures (e-mail, USENET, and the WWW), are asynchronous like the infrastructure used by the listservs considered in this study. This suggests that the literature on voluntary associations and organizations in sociology (e.g. Warner and Hilander, 1964; Warner, 1965; McPherson, 1983, McPherson and Smith-Lovin, 1988; McPherson, 1990; McPherson and Rotolo, 1996) and organization theory (Wilderom and Miner, 1991), is fruitful reference discipline for researchers interested in studying online social structure. Characterization of on-line social structures as associations or organizations also raises questions for future research. The tendency towards low levels of explicit public interaction leads to questions about whether online voluntary associations which do not have many structurally interactive discussions are nonetheless perceived by their members as interactive, and if so, why. Low levels of sporadic communication prompts questions about the processes that might lead to identification, norm
formation, and norm maintenance in long duration, low activity social settings. The flow of people into and out of these structures highlights the dynamic nature of membership in online social structures and leads to questions about the mechanisms and impact of membership changes. Characterizing online social structures as voluntary associations encourages researchers to critically assess the assumptions that have been made in prior work, and from that assessment develop our understanding of "normal" online social structures.

However, while these results illustrate how a dominant metaphor can affect the study of a phenomenon such as on-line social structure, combining metaphors of small groups and voluntary associations is likely to be the most effective strategy for understanding the social environment in emerging communication infrastructures. On one hand, structures based on synchronous communication technology often seem to be 'group-like'. While this may be true within a single on-line session it is often the case that a changing set of individuals interacts over many sessions. Thus, while the operation of particular sessions might be best examined through the lens of the small group metaphor, the dynamics of repeated synchronous on-line social structures could be studied from the perspective of voluntary associations. Likewise, while the structural dynamics of asynchronous on-line social structures, such as listservs, are likely to have commonality with organizations and other macro-social structures, the behaviors of individuals within these structures can be considered from the within the small group framework.

Furthermore, like traditional structures, which make use of both synchronous technologies (e.g. meetings) and asynchronous technologies (e.g. print), on-line social structures are not inherently limited to one type of communication. In addition to considering the nature of on-line structures in different infrastructures, work needs to be done to better model the structural consequences of hybrid infrastructures. Therefore, models that combine features of small group and association
metaphors are likely to provide significant insight into the processes and structures that underlie
the use and impact of emerging communications infrastructures.
References


