Invisible Crowds in Cyberspace:

Mapping the Social Structure of the Usenet[1]

Marc A. Smith

University of California, Los Angeles

SURVEYING SOCIAL CYBERSPACES

The Usenet is a quintessential Internet social phenomenon: it is huge, global, anarchic and rapidly growing. It is also mostly invisible. Although, it is the largest example of a conferencing or discussion group system,[2] the tools generally available to access it only display leaves and branches - chains of messages and responses. None present the trees and forest. With hundreds of thousands of new messages every day, it is impossible to try to read them all to get a sense of the entire place. As a result, an overview of activity in the Usenet has been difficult to assemble and many basic questions about its size, shape, structure and dynamics have gone unanswered. How big is the Usenet? How many people post? Where are they from? When and where do they post? How do groups vary from one another and over time? How many different kinds of groups are there? How many groups successfully thrive and how many die? What do the survivors have that the others lack? How do different social cyberspaces connect and fit together and form a larger ecology?

There is no shortage of questions. But we lack a historical record of the transformations of social cyberspaces[3] just at the point when network interaction media are being widely adopted. Cyberspace is changing the social physics of human life, broadening the size and power of group interaction. But without base-line measures of online activity, we are unable to assess if the groups being selected for study are

typical or if the periods studied are typical for the groups in question. Without a general typology of these spaces there has been no systematic way to contrast research findings and integrate the results of separate research projects. The virtual blackout that cloaks interaction through networks has limited many prior studies to a narrow focus on a specific group over short periods of time (Lewenstein 1992; Pfaffenberger 1996; Phillips 1996), or on individual participants rather than the social space as a whole (Turkle 1996). As a result, attempts to broadly characterize general social processes have been based on limited samples (McLaughlin 1995). While the details of individual experiences and the events occurring in individual groups are important, research should look both at the details of individual groups and at the emergent social structure that grows out of the aggregation of tens of thousands.

Social cyberspaces can be mapped by computer assisted analysis. In this paper I report on the initial results of Netscan, a software tool I designed that gathers an ongoing stream of Usenet messages and maintains a database of information drawn from the header of each message. It then distills measures of activity and relationships in any collection of newsgroups selected for study.[4] This is an initial attempt to survey this emerging social landscape and address some of the limitations of existing tools for exploring them.

Data drawn from group interaction in cyberspace offers some unique opportunities for the study of social organization. One of the unique features of network mediated communication is that almost all interactions leave behind a durable trace: electronic tracks that can provide detailed data about what vast numbers of groups of people do online. Unlike research on face-to-face social relations, online data gathering can be automated and collected over an extended period of time from vast numbers of social spaces, involving millions of people and tens of millions of messages. Longitudinal research on social cyberspaces can illustrate the ways network interaction media, and the social institutions that have emerged in them, have changed over time (Rice 1982). And, because these messages are often created in a casual manner, these data offer insight into the everyday world of many people. Online spaces become self-documenting "natural settings". And while there are a number of possible back channels of communication that limit claims to the completeness of these data, digital artifacts have the advantage of being exact copies of the messages the participants wrote or read themselves.[5] Because online interactions are often archived in their entirety, at least for a short while, the contents of entire collections of messages exchanged by a

group can be examined and studied.[6] Computer generated message headers and accurate copies of message content have even been found to be more reliable sources of data about communications activities than alternative self-reporting surveys (Bikson and Eveland, 1990).

THE USENET

There are a number of reasons to start an investigation into the structure of social cyberspaces with the Usenet. Created in 1979 at the University of North Carolina, as an alternative to services available through the more elite ARPANET, the Usenet initially connected only two computers and handled a few posts or messages a day (Harrison, 1995). Today, it is the third most widely used form of interaction media on the Internet (behind email and the World Wide Web) in terms of users. Growing from fifteen newsgroups, which contained messages that collectively took up fifteen kilobytes per day in 1979, the Usenet now contains more than 14,347 newsgroups carrying two gigabytes of messages per day. On an average day, 20,000 people post 300,000 messages. In the one hundred and fifty days ending November 15th, 1997 1.1 million people[7] posted at least one message each for a total of more than fourteen million unique posts.[8]

The Usenet also has the largest geographic scope of systems of its type, drawing participants from nearly every corner of the globe[9]. By reviewing the email addresses of the people who post to the Usenet we can crudely measure the location of posters around the globe. While not every nation has a large population participating in the Usenet, there are few that are completely absent. Of the 250 officially recognized Internet domains, 238 are reserved for nations and recognized sovereign entities. Of those, only 33 Countries have zero messages in the Usenet.[10] A significant majority of people post from the United States but the system is extremely international, drawing 59% of its participants from outside the U.S. Remarkably, the 15th most common region is anonymous – messages which lack headers that identify author of the post. Information about the author of a Usenet message is stored in the FROM: line of the header. Without much trouble this line can be intentionally modified or damaged so that the author cannot be identified. This is only the most obvious and clumsy form of anonymity practiced in the Usenet. More sophisticated forms of anonymity use properly formatted addresses that point to non-existent people.

Rank	Country	Number of posters	% of Total Usenet Population
1	United States[11]	714,757	40.694%
2	Taiwan	99,150	5.645%
3	Germany	38,848	2.212%
4	Britain	31,657	1.802%
5	Canada	30,116	1.715%
6	Australia	17,801	1.013%
7	Japan	16,322	0.929%
8	Italy	15,203	0.866%
9	France	14,673	0.835%
10	Netherlands	13,251	0.754%
11	Spain	10,013	0.570%
12	Finland	9,178	0.523%
13	Sweden	7,994	0.455%
14	Norway	7,250	0.413%
15	Anonymous (Blank)	6,624	0.377%
16	Poland	6,277	0.357%
17	Russian Federation	6,252	0.356%
18	Denmark	6,015	0.342%
19	Korea (South)	4,708	0.268%
20	Belgium	4,547	0.259%

Table 1. Top twenty national domains appearing in Usenet.

The number of people simply reading the Usenet remains a mystery but it seems safe to assume that more people read the Usenet than actively participate in it. Measurements of readership are difficult to generate because only active participants leave artifacts that can be studied.[12] Newsgroups have many silent observers for every active poster (Terveen, et al. 1997). If we assume a ratio of twenty passive participants to each active one, there could easily be more than twenty million people reading the Usenet at least once a year.[13]

The Usenet helps structure the interactions between these millions of participants by organizing every message into a branching tree structure. As seen in Figure 1., each newsgroup is located within one of a number of "hierarchies" which broadly categorize all of the topics covered in the Usenet. There are 143 widely distributed hierarchies, but a group of eight (alt, comp, misc, news, rec, sci, soc, and talk) are the historical core of the Usenet (Salzenberg 1992). Newsgroups are named in loose accord with a number of conventions that describe their topic and goals. Groups start with one of the hierarchy names and then add words separated by periods that increasingly narrow the scope of the group, for example, "*comp.lang.perl.misc*", "*alt.support.diabetes.kids*", and "*misc.kids.pregnancy*".

Each newsgroup contains a collection of posts that are organized into "threads" - chains of posts linked like paper clips to one another. Any post can be "crossposted", a practice that places a copy of the message in a number of separate newsgroups.[14] Each post contains a "header" that automatically records a range of information about the post, including the email address; user name and organizational affiliation of each message's author; the subject of the message; a number that identifies each message uniquely; the date and time the message was posted by the author; a list of all the newsgroups it is intended to be distributed to; and information that locates the message within a chain of other messages.[15]
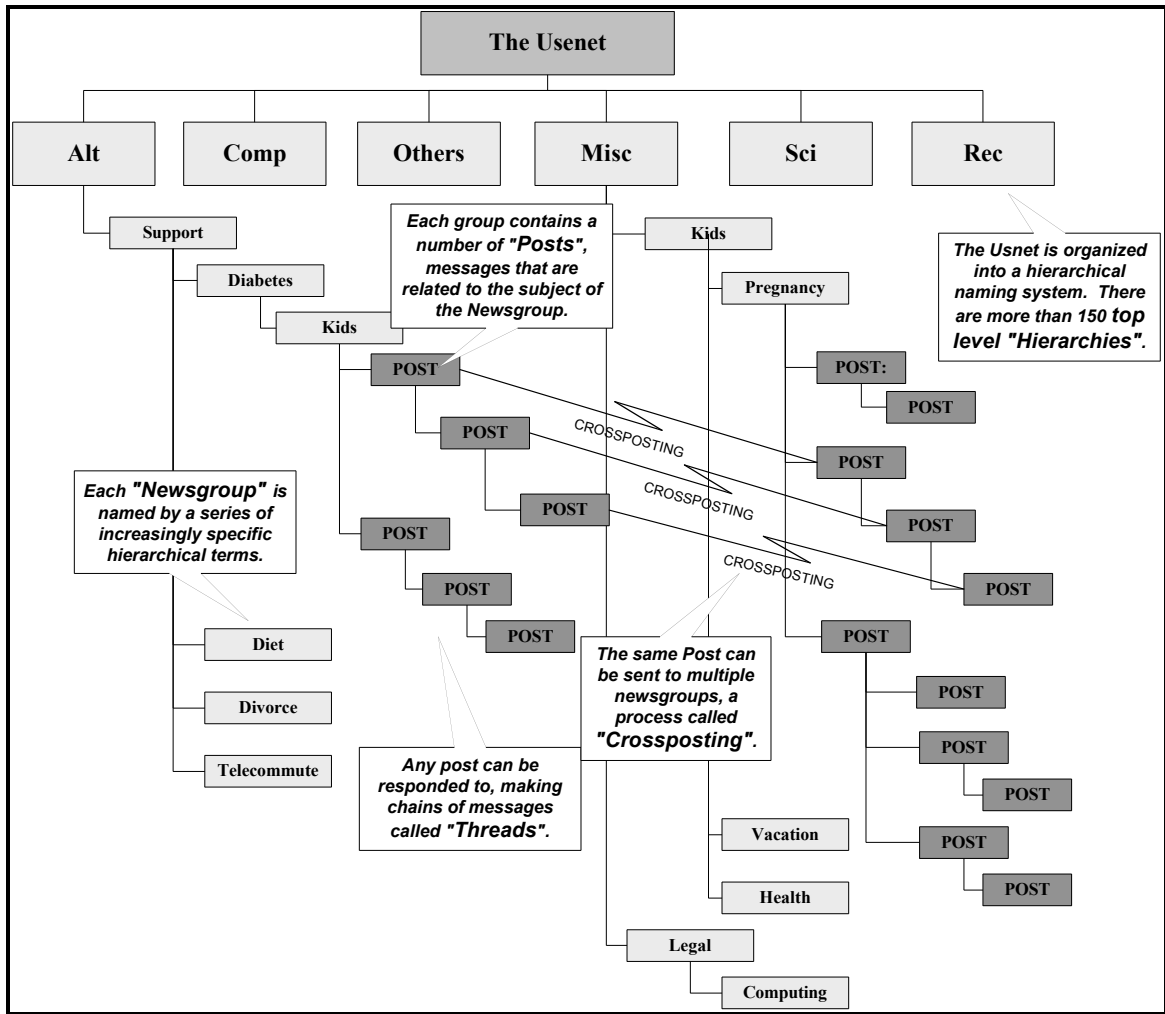

Figure 1. Schematic Representation of the Usenet

More than its size, scope or structure, it is the Usenet's technical and social organization that makes it of particular interest. The Usenet can be divided into two main components: the technical and physical infrastructure, the computers, wires, phone lines, and software; and the social and technical structure of its content, the system of interconnected newsgroups, threads, and messages. Both the technical and social structure of the Usenet, analogous to the phone system and the conversations that take place through it, are decentralized and organized as a commons. These qualities play an important role in shaping the kinds of groups that form within it.

In contrast with systems run and managed by a single organization, the Usenet is not a commercial product and has grown without central planning or control. No single individual, group or cabal is in complete or even dominant control. None of the more than 300,000 computers around the planet that act as Usenet hosts is the central point of the entire system. Each system owner, from information system managers at large corporations and universities to individual hobbyists, decides what part of the feed to accept and what to pass on and to whom. Usenet members agree simply to accept and pass along the "feed" - the collection of messages currently being shuttled between all the Usenet "hosts" in the world. The Usenet is like an informal bucket brigade, as shown in Figure 2. Someone who wants to carry the contents of the Usenet only has to convince a person who already receives it to pass it along to them. The technical structure of the Usenet shares many of these qualities with the Internet itself, both are social and technical achievements that exist only because of widespread informal agreements to minimally cooperate with *local* partners.
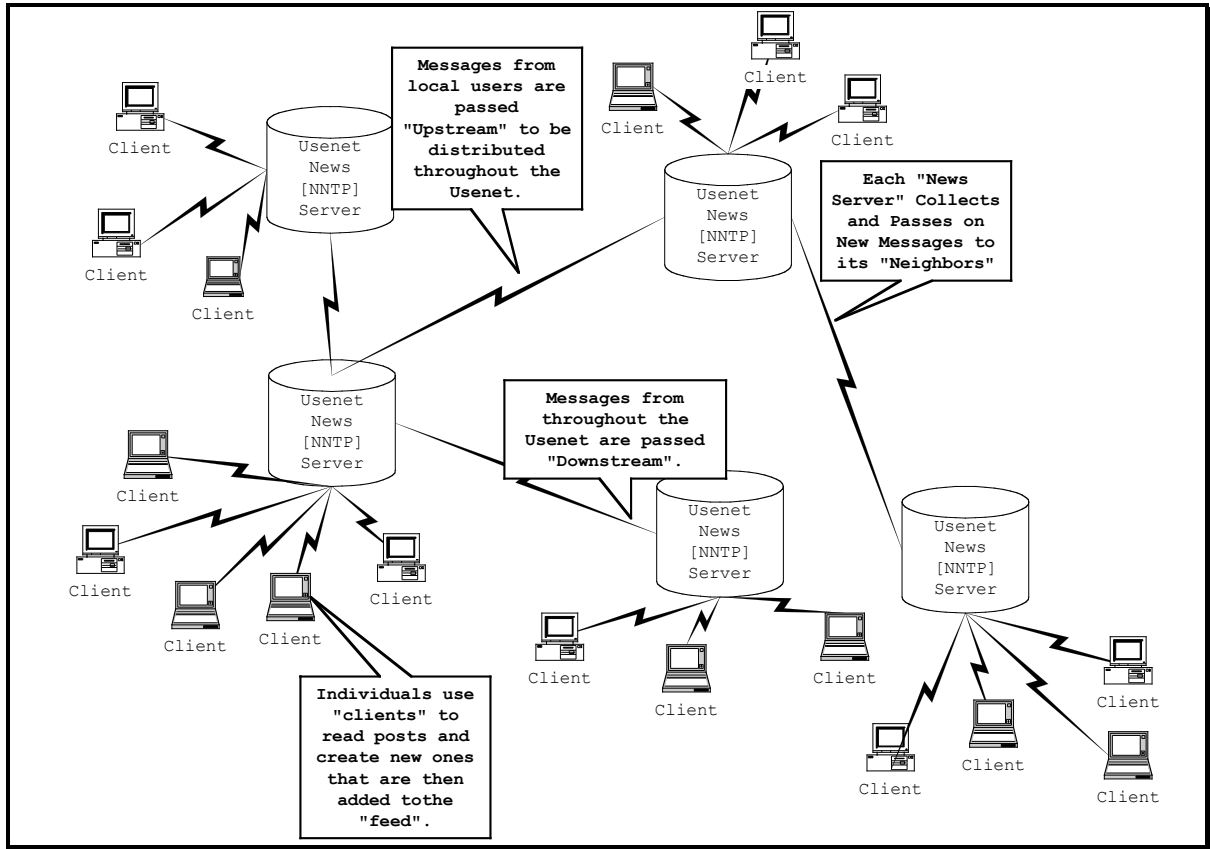
Figure 2. Usenet Technical Organization

This technical infrastructure is the foundation for the formation of the Usenet's social structure:

tens of thousands of social spaces or "newsgroups" devoted to a huge range of interests and groups. Each

newsgroup is potentially a place where people can collectively produce something that none could have

created by themselves. Some are practical and convenient sources of technical support and information.

Others provide entertainment, a sense of connection, membership and mutual support. From around the

planet, expectant mothers swap stories about their experiences in *misc.kids.pregnancy*, programmers with

problems they can't solve on their own find guidance in *comp.lang.perl.misc*, video game fans offer tips to

one another in *comp.sys.ibm.pc.games.strategic*. Newsgroups are the source of public goods, the resources

produced in them are available to any one even if they do not pay for or contribute to their creation.

However, the potential for creating collective projects in these network commons does not

guarantee their success. As Mancur Olson noted, "if the members of some group have a common interest

or objective, and if they would all be better off if that objective were achieved, it [does not necessarily

follow] that the individuals in that group act to achieve that objective." (Olson 1965) Common sense would suggest that none of the groups created through network interaction media should succeed. Nearly anonymous people from around the world with no prior introduction independently request or contribute time and expertise and freely give the result away to anyone interested without payment or coercion. Such a fanciful social organization would seem doomed to failure.

This challenge is even greater in the Usenet, since many of the resources groups usually rely on to help maintain themselves are oddly transformed or missing. The distributed technical architecture of the Usenet has far-reaching implications for the kinds of social organizations that form within it. Most conferencing systems (for example, message boards in America Online) and alternative network interaction media (like email, chat, and muds) are centralized and "owned" by a single person or institution. Because they occupy a central location in their system, owners of these spaces have the ability to limit people's access and filter their contributions. In the Usenet few of the spaces are "owned" in this way.[16] As a result, it is almost impossible to stop someone from posting a message to a newsgroup or reading the messages that are already there. No one can be banned and no one can be gagged.[17] Any and every user of the Usenet potentially wields an equal and large amount of control over the entire system. This makes the Usenet about as pure a case of social anarchy as can be found in the real world. The lack of boundaries ensures that newsgroups remain a kind of commons where their contents are publicly available to any and all who are interested (Kollock and Smith 1996).

Given the challenges, it is not surprising that the tens of thousands of newsgroups that make up the Usenet are not uniformly pleasant, useful, or even entertaining. In many cases a newsgroup is a barren or cacophonous space that may yield less value than the effort it takes to be there. Many groups seem to be dedicated to non-cooperative interaction: people fill them with challenges, insults, irrelevant advertisements and lures for scams. On balance there may even be more useless material than good. The focus on failures has led to a common cliché in Usenet discussions -- "Imminent death of the Usenet predicted" (Raymond 1993). But after seventeen years of repeated announcements of its impending demise, the reality is that the Usenet is a robust social institution that permanently teeters on the brink of chaos. The Usenet displays an impressive robustness[18], sustaining useful social interaction in the face of (or perhaps because of) anarchic organization, rapid growth[19], and intentional attacks.[20] The undeniable fact is that many people find islands

of quality interaction and access valuable resources through media like the Usenet.[21] The question is not why many of these collective projects fail, but why as many that succeed do so.

VARIATION IN SOCIAL CYBERSPACES

But how many successes are there? The question is difficult to answer; what constitutes "successful"? An alternative and more tractable question is what kinds of interaction patterns are present in each newsgroup? What immediately becomes clear is that social cyberspaces are not all the same. The following takes each logical level of the Usenet, hierarchies, newsgroups, posts, posters, and crossposting in turn looking at the range of variation across the system as a whole.

Hierarchies

Each hierarchy varies in terms of the number of groups it contains, the number of messages those groups receive and the number of people who contribute those messages.

Rank	Hierarchy	Description	# Groups
1	alt:	"Alternative" groups	5339
2	comp:	Computer related	884
3	rec:	Recreational	677
4	clari:	Clarinet commercial news service	584
5	uiuc:	University of Illinois, Urbana Champagne	406
6	microsoft:	Microsoft	355
7	fj:	Japanese Kanji Groups	353
8	ucb:	University of California, Berkeley	335
9	ucla:	University of California, Los Angeles	328
10	de:	Germany	313
11	tw:	Taiwan	279
12	soc:	Social issues	258
13	bit:	Bitnet gateway newsgroups	230
14	sci:	Science	203
15	zer:	More German Groups	191
16	uw:	University of Washington	169
17	sfnet:	Finland	151
18	aus:	Australia	132
19	misc:	Miscellaneous	131
20	relcom:	Russia	131

Table 2. Top twenty hierarchies present in the Usenet

As seen in table 2, of all the hierarchies, "alt" is by far the largest. The "alt" hierarchy is dedicated to a range of subjects that fall outside the range covered by the other categories. "Alt" possesses 29% of all newsgroups, 22% of all the postings, and 24% of Usenet participants. This may be because the "alt" newsgroups are a political sub-continent of the Usenet, governed by a different and less restrictive process for the creation of new newsgroups.[22] This means that the most active area of the Usenet is not covered by the same political regime that rules the others. It does not mean that this activity equates to quality, value or user satisfaction. But it does suggest that the difference in social regulation plays a role in the difference in activity.

GROUPS

It is not easy to measure the exact number of newsgroups. The Usenet's architecture, essentially an amorphous distributed database, causes it to look differently depending on the news server from which it is viewed, making claims to the completeness of any study questionable. Since every local news server's feed is partial (almost no single news server takes every newsgroup), a truly complete picture of the Usenet may be impossible to generate.[23] Netscan studies show more than 79,000 newsgroups exist worldwide (although many may be only locally distributed). Of these, my study collected data on the 14,347 groups carried in the UCLA news server's feed.[24]

POSTS

Newsgroups do not all receive the same number of messages. A full fifth of the newsgroups studied were entirely empty.[25] Many newsgroups (42%) held fewer than one hundred messages in the ten-week period studied. A significant mid-range of newsgroups (23%) contains between one hundred and one thousand messages over this period. The remaining messages are distributed throughout a smaller set (7%) of "super newsgroups" that contain more than one-thousand posts. At the very top of the scale are a tiny collection

of one-hundred-and-sixty-three newsgroups (~1%) that range from more than a thousand to as many as two-hundred-and-fifty-thousand messages over the entire period. In general, the Usenet is mostly filled with small newsgroups with a moderate amount of message traffic.

The three largest, and seven of the top ten newsgroups in terms of numbers of messages, are devoted to job announcements (Table 3.). While these job related groups, many with a specific regional focus, are the largest throughout the entire Usenet, they have dramatically fewer participants than smaller groups dedicated to other topics. Many of the messages in these groups are posted by commercial organizations soliciting job applications for technology related positions and are more like broadcast channels than social interaction spaces.

Rank	Group	#Posts	#Posters	Poster to Post Ratio	Xpost Degree	#Articles Xposted
1	Misc.jobs.offered	232,612	5,184	0.022	793	217,845
2	Biz.jobs.offered	217,472	3,227	0.015	648	202,977
3	Ba.jobs.offered	210,562	2,249	0.011	433	169,611
4	Misc.jobs.contract	98,803	2,353	0.024	481	95,731
5	alt.jobs	76,605	1,438	0.019	537	76,600
6	ba.jobs.contract	29,077	1,139	0.039	246	27,850
7	news.newusers.questions	27,332	12,012	0.439	827	6,470
8	comp.sys.ibm.pc.hardware.video	25,293	5,576	0.220	374	8,579
9	tx.jobs	20,741	931	0.045	335	19,394
10	comp.sys.ibm.pc.games.strategic	20,482	5,055	0.247	221	7,987

Table 3. Top ten newsgroups in terms of message volume November 1, 1996 - January 18, 1997

POSTERS

The groups of people that contribute to Usenet newsgroups are usually very small, mostly fewer than fifty people. About a quarter of all newsgroups attract between fifty and five-hundred people. These groups may be the most productive and stable of all newsgroups. There is some indication that face-to-face groups can maintain self-organized cooperative relationships fairly easily when group sizes remain small, usually below one-hundred-and-fifty people (Orbell and Dawes 1981; Messick and Brewer 1983). One effect of the ways network mediated communication alter the economies of interaction may be to double or triple this maximum level, allowing cooperative groups to expand to include as many as five or six hundred active people and many thousand more passive participants. A few newsgroups are huge, drawing thousands of different participants. Further research may be able to determine if these larger groups are more or less stable and cooperative than smaller groups.

On average, there are four messages contributed to the Usenet for each poster. But posting behavior is far from average. Posting behavior resembles a Poisson distribution, starting at a peak of 42% of all posters who posted only a single message in the ten week period studied and falling off steeply with a tiny fraction posting more than 200 messages. The broadest pattern of activity is sparse and sporadic, 96% posters posted fewer than thirty messages in the period. Remarkably, the 99.55% people who posted fewer than 200 posts are responsible for only 63% of all messages. In contrast only 0.04% of all posters posted

more than 200 posts. But this group created 37% of all the messages contributed in the study period. This small group of active posters is probably composed of non-human posters (i.e. control messages and spam).

The most popular newsgroups, where the most people contribute messages, focus on topics related to the Usenet itself, employment opportunities and aspects of personal computer use. These groups are good examples of the way the Usenet is used to sustain mutual support groups that can draw upon a vast population. But not all newsgroups, even those that contain large numbers of messages, are places where social interaction occurs. Some are barren, completely abandoned by everyone. Others can be kinds of broadcast channels, cleared of interaction in order to let announcements stand out from the clutter.

One rough measure of the quality of interaction in a newsgroup is its poster-to-post ratio. Some newsgroups are characterized by a high poster-to-post ratio, that is, one in which many different people write most of the messages. A newsgroup with a lower measure indicates that a few participants are contributing a disproportionate number of messages to the newsgroup. A measure closer to one indicates a lack of turn-taking social interaction since each person writes only one message. This means that no one responds to a response, a sign of the turn-taking structure indicative of direct social interaction (Sacks, et al. 1974). Yet newsgroups with high poster-to-post ratios are not necessarily barren and useless. Newsgroups set aside for announcements only, where discussions are discouraged to provide an uncluttered place for messages of higher priority, can have a high poster-to-post ratio.

A poster-to-post ratio measure closer to zero indicates that a newsgroup has very few active participants who contribute many messages. Most of the "clari.*" newsgroups have a very low poster-to-post ratio. These groups are part of a commercial news service called Clarinet that piggybacks on the Usenet to distribute wire service news stories and financial information. In "clari.*" groups a single poster, the Clarinet news service, alone posts thousands of messages.
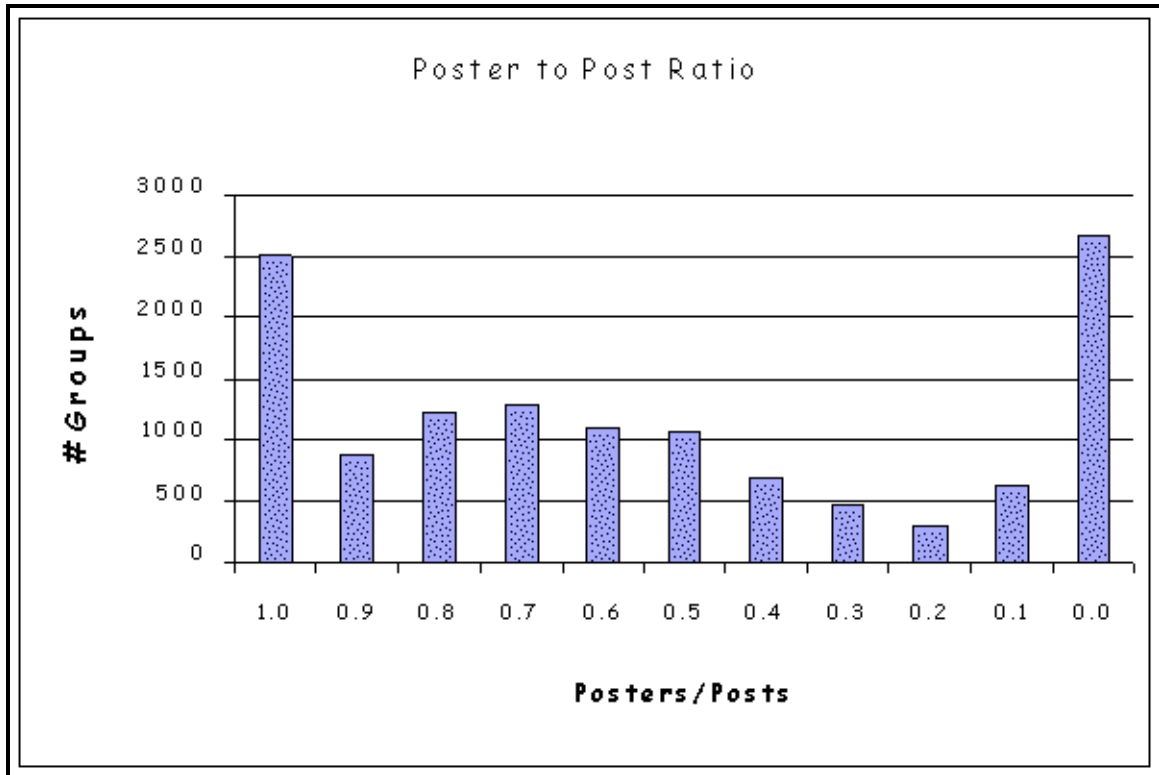
Figure 3. Distribution of Poster-to-post ratio for all newsgroups, November 1, 1996 to January 18, 1997.

Extremely high or low poster-to-post ratios indicate a lack of interaction between participants in a newsgroup. A poster-to-post ratio in the middle range is not a guarantee of social interaction (a number of people may post more than once and still not be talking to each other) but measures at either extreme are solid indications of a lack of turn-taking style interaction. While newsgroups are generally thought of as interactive places, only 60% of all newsgroups have poster-to-post ratios that indicate conditions that allow for interaction.

Rank	Group	#Posts	#Posters	Poster to Post Ratio	Xpost Degree	Xpost Volume
1	news.newusers.questions[26]	27332	12012	0.439	827	6470
2	comp.os.ms-windows.win95.misc	16875	7189	0.426	516	9879
3	comp.sys.ibm.pc.hardware.video	25293	5576	0.220	374	8579
4	misc.jobs.offered	232612	5184	0.022	793	217845
5	comp.sys.ibm.pc.games.strategic	20482	5055	0.247	221	7987
6	comp.sys.ibm.pc.hardware.chips	12005	4394	0.366	303	6021
7	comp.os.ms-windows.nt.misc	10337	4377	0.423	341	5553
8	comp.sys.ibm.pc.games.action	11525	4329	0.376	265	6931
9	comp.os.ms-windows.win95.setup	7792	4202	0.539	286	5305

| 10 | comp.sys.ibm.pc.games.rpg | 12590 | 4201 | 0.334 | 243 | 6113 |

Table 4. Top ten newsgroups in terms of poster population, November 1, 1996 - January 18, 1997.

Of the ten largest newsgroups all, with the exception of *misc.jobs.offered*, have poster-to-post

ratios in the middle range, an indication that these groups are characterized by many participants each

contributing a number of messages. This reflects the general purpose of these newsgroups where people

ask and answer technical questions pertaining to the configuration and operation of personal computers and

popular games played on them.

Daily and Weekly Cycles

On an average day, 300,000 messages are posted by 18,000 posters. Messages arrive at the rate of about

thirty-five-hundred messages an hour, written on average by twelve-hundred posters. At its peak, 400,000

messages are created per day by forty-four-thousand posters. Messages are not contributed at a steady rate

throughout the day. In fact, the Usenet has a weekly cycle of activity that builds during the workweek and

falls off over the weekends, suggesting that many people access the Usenet from their workplaces. This

may challenge the belief that recent network growth has been predominantly driven by home consumer use.
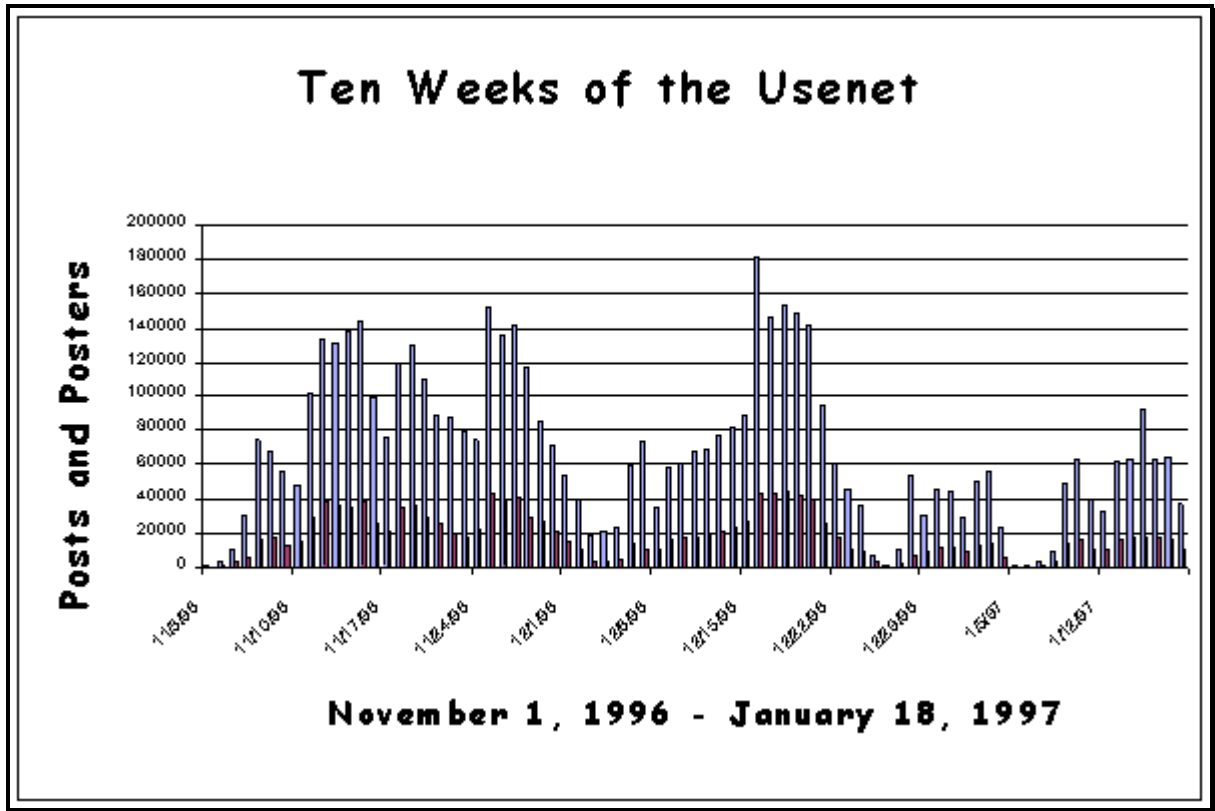
Figure 4. Daily rates of messages and participants in the Usenet, November 1, 1996 to January 18, 1997.

As seen in Figure 4, the fairly stable weekly pattern found in November dissolves in December, where usage rises continuously until just before Christmas Day, which was one of the lowest activity days in the study. In contrast to Christmas, which is celebrated more globally, the relatively normal activity on Thanksgiving may be an indication that activity outside the United States, where Thanksgiving is not as widely celebrated, is a significant portion of overall Usenet activity.
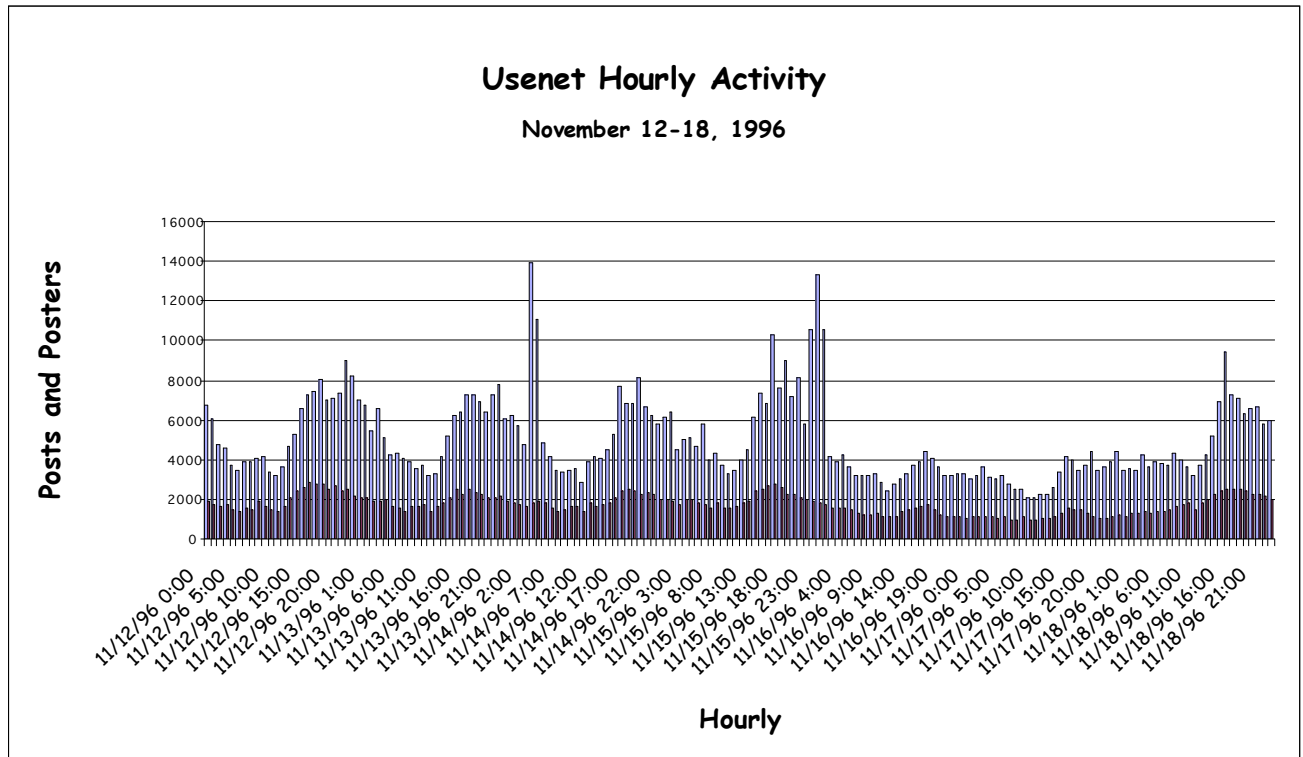
Usenet Hourly Activity

November 12-18, 1996

Figure 5. Hourly rates of messages and participants in the Usenet, November 1, 1996 to January 18, 1997.

Each day, as illustrated by Figure 5, activity in the Usenet begins to rise at about four o'clock in the afternoon Greenwich Mean Time[27]. This is about eleven o'clock in the morning on the East coast of the United States, eight o'clock in the morning on the West coast. Each day the lowest points of activity occur during the European day and North American night. But even at its low points, the Usenet is always active, an indication of both the global scope of the participants in the Usenet and the way asynchronous communication allows people to interact at any time of the day or night.

Crossposting Connections between Newsgroups

Through the practice of crossposting, newsgroups are densely interconnected with one another, forming large neighborhoods of interrelated topics.[28] These connections can be measured in two ways. First, each newsgroup is connected to a certain number of other newsgroups. This is the crossposting degree. A related but separate measure is the crossposting volume, the count of the number of posts the groups shares

through crossposting with other newsgroups. A newsgroup could be connected to many other groups but share only a small portion of its messages with them, or a newsgroup can be connected to only a few newsgroups, but share many of its messages. Most interfaces to the Usenet present each newsgroup as if it were distinct and isolated. But very few newsgroups are in fact islands, only seven-hundred-and-ninety-eight (6%) active newsgroups are not connected to any other newsgroups at all. While most newsgroups are connected to only a few others they are often directly connected to one of the core newsgroups. A core newsgroup is one to which most other newsgroups are connected. Because of the presence of these core groups, no newsgroup is more than a few "steps" apart from any other. On average, a newsgroup is connected to fifty other newsgroups. At the top end of the spectrum, five-hundred-and-eighty-eight (4%) are connected to more than two-hundred other newsgroups. While only 15% of all groups are connected to more than one-hundred other groups through crossposting, these massively crossposted newsgroups are the core of the Usenet, containing 69% of all the messages posted during the study.

This dense level of interconnection gives the Usenet the ability to act as a powerful social information switch. Questions that appear in one newsgroup are likely to be seen by someone who has a connection with a more appropriate newsgroup, who then forwards the message or redirects the questioner to a proper newsgroup.

The top ten most crossposted newsgroups are not necessarily the most active or populous newsgroups. But their high levels of interconnection make these groups the crossroads of the Usenet. At the same time, high levels of crossposting are indications that a newsgroup lacks a clear boundary, possibly indicating the lack of focused content and a stable population. More distinct boundaries may be necessary for the emergence of social ties that characterize groups that are closer to what is commonly understood to be a "community".

Rank	Group	#Posts	#Posters	Poster to Post Ratio	Xpost Degree	#Articles Xposted
1	news.answers	4,350	670	0.154	992	4,346
2	Alt.forsale	4,745	1,881	0.396	960	4,727
3	alt.business	10,236	1,533	0.150	835	10,208
4	news.newusers.questions	27,332	12,012	0.439	827	6,470
5	misc.jobs.offered	232,612	5,184	0.022	793	217,845
6	alt.retromod	182	32	0.176	762	182
7	alt.sex	2,128	1,216	0.571	724	2,087

8	alt.business.misc	11,699	1,525	0.130	720	11,671
9	misc.entrepreneurs	14,760	2,485	0.168	700	12,016
10	misc.misc	7,957	1,536	0.193	677	7,255

Table 5. Top ten newsgroups in terms of crossposting degree, November 1, 1996 - January 18, 1997.

In practice, when posters crosspost many messages to another newsgroup they can effectively merge into the same group. Readers and participants in each closely crossposted newsgroup are effectively members of a single larger meta-newsgroup. There is good reason to believe that there are far fewer meta-groups than distinctly named newsgroups since 58% of newsgroups crosspost more than 50% of the messages they contain. Remarkably, in one-third of all newsgroups, 100% of the messages are crossposted to (or from) other groups.

Using crossposting records, it is possible generate network maps of the interconnections between Usenet newsgroups.[29] Crossposting patterns can be more intuitively represented as a network map (Krackhart, et al. 1994; Becker, et al. 1995; Cox and Eick 1995; Krebs 1996). The patterns of a newsgroup's connections to others may indicate a great deal about the newsgroup (Rice 1995). It is beyond the scope of the current project, but a map of the interconnections of all the newsgroups is entirely feasible.
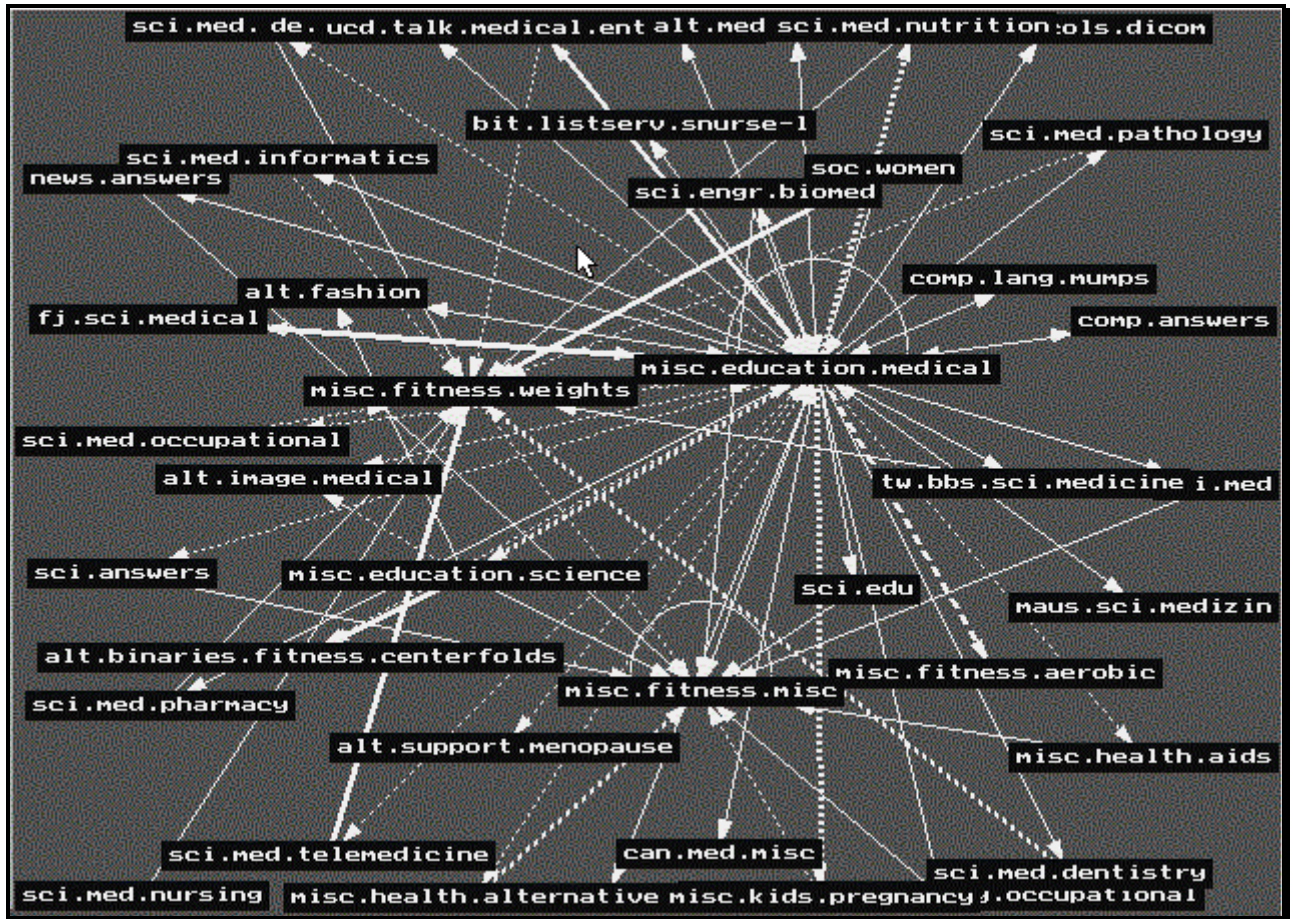
Figure 6. The misc.education.medical crossposting neighborhood, November 1, 1996 to January 18, 1997.
These maps can highlight some unexpected connections between newsgroups. For example in

Figure 11, an unexpected connection between alt.fashion and misc.education.medical is created by

recurrent discussions of the health effects of dieting. These maps illustrate the ways cultural topics and

interests are clustered together and can be useful guides to direct observation of related groups.

DISCUSSION

These initial results illustrate some important features of the Usenet. Within the sample I examine, the

average newsgroup has about one hundred messages a week, contributed by fewer than fifty different

people. Nearly all of the messages in a majority of newsgroups are crossposted to other newsgroups, and

each group is connected to fifty other groups on average. On an average day, eighteen thousand people

contribute sixty-seven-thousand messages to the Usenet. In an average hour, thirty-five-hundred messages

are written by twelve-hundred people. The high rates of interconnection between newsgroups through

crossposting means that there are fewer newsgroups than there are newsgroup names. A majority of newsgroups crosspost more than 50% of their messages, making them interconnected to the point that they agglomerate into larger meta-clusters. Mapping these clusters is a key direction for future research. The highly interconnected nature of the Usenet may be the quality that accounts for its extraordinarily robustness.

Future Directions

These measures begin to answer some basic and important questions about the ways human groups work when computers and networks mediate them. A variety of improvements in the Netscan software promises additional information about the social and technical structure of the Usenet. Future study may be able to show how distributed cooperation works in network interaction media. The existing database can be analyzed further to reveal information about the patterns of posting behavior, the connections between individuals and between newsgroups, and the life cycles of individual newsgroups. Other forms of network interaction media including other conferencing systems, email (either from the perspective of an individual or a mailing list), and some forms of real time chat systems and MUDS, could be studied, allowing comparative analysis of different network interaction media. Two areas of research, on the presence and size of core groups of posters and the structure and length of the threads in newsgroups stand out in particular.

Some newsgroups are populated by a core of dedicated participants who contribute much of the value found in the group.[30] A core group of posters can act as a means of socialization, ensuring that group experiences and lessons are conveyed to the next generations of participants. The presence or absence of a core group who produce a significant and disproportionate amount of the participation may explain why some newsgroups are more ordered and productive than others. Core posters, who make up only a few percent of most newsgroup's population, post more than twenty times the average amount of messages, target their messages more finely, and are more likely to be responded to (Terveen et al. 1997).

Another measure that captures a newsgroup's quality of interaction is the thread to post ratio. Newsgroups in which the dominant form of interaction is the contribution of announcements that receive no commentary will have a very high thread to post ratio: each thread will contain only one post. Newsgroups in which the dominant form of interaction is a fairly straightforward pattern of questions asked and answered will have thread to post ratios near 0.5, each thread will have one question and one answer. If the measure is lower, the newsgroup is more likely to be characterized by long threads which may be an indicator of discussions characterized by disagreement, divergent opinions or ambiguous subjects that can not be neatly characterized. Initial results show that 11 million of the 14 million messages posted in the 150 days ending 17 November 1997 are "thread heads" – initial posts that indicate the beginning of a new chain of conversation. This means that only about 3 million (21%) messages receive any replies at all.

Content analysis of message bodies could lead to studies that map the diffusion of topics through the Usenet and other communication media. Such data could assist studies of informal communication networks and the transmission of folk beliefs (Shibutani 1958) as well as the development of academic disciplines (Bayer et al. 1990). Studying message content could allow for the measurement of social problems like mass mailings, or "Spam", that clutter many newsgroups.

Ethical Issues

Digital artifacts are incomplete and potentially dangerous. There are reasons to be cautious about research using data collected from network interaction media. These data refine and extend the means of surveillance that are already a disturbing trend in such systems. For all of the potential benefits, using data of this sort requires consideration of the possible impact of such research on the people studied. Data that maps the activities of thousands of individuals around the world raises serious issues about privacy, property and the responsibilities of researchers to the members of the online groups they study. The information available through machine analysis of the artifacts of network interaction can uncover social spaces, subjecting them to a kind of panoptic surveillance (Poster 1990).

In effect, such research produces maps that reveal in great detail a social space that has previously been cloaked. Is the creation of such a map an invasion of privacy? The answer depends largely on the nature and status of the Usenet and other social cyberspaces. Are these public spaces, like city streets or plazas, where we have lower expectations and protections of our privacy? Or are these private or semi-private spaces where people have an expectation of privacy? Even if Usenet is a "public" space or a commons, are all actions in it subject to aggregation and analysis? If many people have a high expectation of privacy, do those expectations deserve respect regardless of the technical realities? Or do new technical tools change and shift the nature of this space, forcing participants to recognize a change in its character?

As networks and computers insinuate themselves into more and more aspects of everyday life, the use and interpretation of these records raise a range of ethical and legal issues concerning the protection of individual privacy (Clarke 1994) [31]. The goal must be responsible and ethical research practices that allow researchers to derive much of the value of this information, without eroding the privacy of participants. But researchers have a responsibility to consider the impact of their work. It can be argued that some information should not be gathered even if it is easy and possible to do so. King argues for balancing the amount of deception and disclosure practiced in the process of data collection with the extent to which the data is anonymized (King 1996).[32] He also notes a perverse effect of online studies: research that violates an online group's sense of privacy may leave "scorched earth" behind for prospective participants and future researchers as participants seek more private online spaces to carry out their group's business or simply scatter under the scrutiny of researchers. The bright light of social science research can create an unpleasant glare for participants drawn to a dimly lit online space. And, as Reid notes, this effect can occur even when researchers comport themselves responsibly but draw other researchers to the space they studied (Reid 1996). The light cast by researchers can also act as a beacon for others, making the space all but useless for its participants. Online researchers have the potential of becoming social locusts, descending on online spaces and rendering them barren.

The stakes are even higher when research reveals information that can be linked to specific individuals, even more so if an individual's diverse and unrelated activities can be drawn together into a unified dossier of all their activity (Clarke 1994). Researchers may have a responsibility to anonymize both the identities of the participants in an online group and the location of the online space itself.

As with many issues related to computer networks, the issue may already be moot. Regardless of whether researchers include personal identification or identify specific newsgroups, anyone can now generate similar reports on individuals and newsgroups with DEC's Alta Vista service (http://altavista.digital.com/) or with Deja News (http://www.dejanews.com/). Alta Vista, like a half a dozen similar services, allows users to enter a person's name or email address and get a list of all the messages that person contributed to all Usenet newsgroups since the service started collecting data in March of 1996. Deja News[33] goes a step further, explicitly offering a "Poster Profile" report that lists the number of times the person selected posted to the Usenet over a period of time, the percentage of postings that were responded to as well as a breakdown of each newsgroup the person posted to and the number of messages the person posted there. This profile can be seen as a very informative aid in the evaluation of the merit of a poster's message or as an invasion of privacy. Regardless, these services do create a dramatic change in the balance between self-exposure and self-disclosure.

Groups may find that there are benefits to operating within a system that is surveilled by network services like AltaVista, DejaNews, and Netscan. The data collected about group activity may help groups that are studied to see themselves. The data collected about group dynamics can be reflected in the user interface to these spaces providing information about the overall population and group sizes, level of interaction and the internal connections and structure of the groups that are missing from existing systems. Offering this information directly to the users of these media may return to the participants some of the information that is lost when communicating through network interaction media.

References

Bayer, A. E., J. C. Smart, & G. W. McLaughlin (1990). Mapping intellectual structure of a scientific subfield through author cocitations. Journal of the American Society for Information Science, 41(6):444-452.

Becker, Richard A., Stephen G. Eick and Allan R. Wilks, "Visualizing Network Data," IEEE Transactions on Visualization and Computer Graphics, Vol. 1, No. 1, pages 16-21, March 1995.

Bikson, T. and J. D. Eveland 1990. The Interplay of work group structures and computer support. In J. Galegher, R. Kraut, and C. Egido (Eds.), *Intellectual teamwork: Social and Technical Bases of Cooperative Work* (pp. 245-290). Hillsdale, New Jersey: Lawrence Earlbaum.

Brewer, Robert S. and Philip M. Johnson, Collaborative Classification and Evaluation of Usenet", Collaborative Software Development Laboratory, Department of Information and Computer Sciences, University of Hawaii, Honolulu, HI. Unpublished manuscript. 1996

Clarke, Roger. 'The Digital Persona and Its Application to Data Surveillance' The Information Society 10,2 (June 1994).

Cox, Kenneth C. and Stephen G. Eick, "3D Displays of Internet Traffic," IEEE Symposium on Information Visualization, eds. Nahum Gershonand and Stephen Eick, pages 129-131, IEEE Computer Society Press, October 1995.

Cox, Kenneth C. et al, "3D Geographic Network Displays," Sigmod Record, Volume 24, Number 4, December 1996.

Damaschek Marc Gauging Similarity with n-Grams: Language Independent Categorization of Text. Science Vol. 267 10 February 1995.

Danowski, J. (1982) Computer-mediated communication: A network-based content analysis using a CBBS conference. In M. Burgoon (Ed.), *Communication Yearbook* (Vol. 6, pp. 905-924). Beverly Hills, CA: Sage.

Danowski, J. (1988) Organizational infographics and automated auditing: Using computers to unobtrusively analyze communication. In G. Goldhaber & G. Barnett (Eds.) *Handbook of organizational communication* (pp. 385-434). Norwood, NJ: Ablex.

Eick, Stephen G. "Aspects of Network Visualization," Computer Graphics and Applications, Vol. 16, No.2, pages. 69-72, March 1996.

Harrison, Mark, The Usenet Handbook: A User's Guide to NetNews, O'Reilly and Associates, 1995.

Hess, Charlotte Untangling the Web: The Internet as a Commons, 1996 Paper Presentation at the
"Reinventing the Commons" Workshop, Transnational Institute, Bonn, Germany, November 4-5,
1995.

Hill, Will and Loren Terveen, "Using Frequency-of-mention in public conversations for social filtering",
unpublished manuscript. 1996. http://weblab.research.att.com//phoaks.cscw96.ps

Hiltz, SR, Turoff M. 1993. *The Network Nation: Human Communication via Computer*. Cambridge, MA:
MIT Press.

Jones, Steve G. 1995. Understanding Community in the Information Age. In *CyberSociety: Computer-
Mediated Communication and Community,* ed. SG Jones, pp-10-35. Thousand Oaks, CA: Sage.

Kiesler, Sara "The Hidden Messages in Computer Networks," Harvard Business Review, January-February
1986.

King, Storm. "Researching Internet Communities: Proposed Ethical Guidelines for the Reporting of
Results". *The Information Society*, 12(2), 119-128.

Kling, Rob. Social Relationships in Electronic Forums: Hangouts, Salons, Workplaces and Communities.
Computerization and Controversy: Value Conflicts and Social Choices (2nd Ed.) by Rob Kling.
San Diego, Academic Press, 1996.

Kollock, P, Smith MA, 1996. Managing the Virtual Commons: Cooperation and conflict in computer
communities. In Computer Mediated Communication: Linguistic, Social and Cross-Cultural
Perspectives, Susan Herring, eds. John Benjamins, 1996.

Korenman, Joan and Nancy Wyatt, Group Dynamics in an Email Forum, in Computer Mediated
Communication: Linguistic, Social and Cross-Cultural Perspectives, Susan Herring, eds. John
Benjamins, 1996.

Krackhardt, D., Blythe, J., and McGrath, C. 1994. "KrackPlot 3.0: An Improved Network Drawing
Program" Connections, Vol. 17(2):53-55 (Dec).

Krebs, Valdis. 1996. "Visualizing Human Networks", Release 1.0, 12 February 1996.

Larson, Ray R. Bibliometrics of the World Wide Web: An Exploratory Analysis of the Intellectual
Structure of Cyberspace, Electronic Document: http://sherlock.berkeley.edu/asis96/asis96.html

Lewenstein, Bruce V. The Changing Culture of Research: Processes of Knowledge Transfer. OTA 28
December 1992.

Marsden Peter V. Network Data and Measurement. Annual Review of Sociology 1990 16:435-463.

McLaughlin, Margaret, et al. 1995. Standards of Conduct on Usenet. In *CyberSociety: Computer-
Mediated Communication and Community,* ed. SG Jones, pp-10-35. Thousand Oaks, CA: Sage.

Messick, David M., and Marilynn B. Brewer. 1983. "Solving Social Dilemmas." Pp. 11-44 in Review of
Personality and Social Psycology (Vol. 4), edited by L. Wheeler and P. Shaver. Beverly Hills, CA:
Sage.

Olson. Mancur, Jr. 1965. The Logic of Collective Acton. Cambridge, MA: Harvard University Press.

Orbell, John, and Robyn Dawes. 1981. "Social Dilemmas." Pp. 37-65 in Progress in Applied Social
Psychology (Vol. 1), edited by G.M. Stephenson and J.M. Davis. New York: Wiley and Sons.

Pfaffenberger, Bryan. 1996. "If I Want It, It's OK": Usenet and the (Outer) Limits of Free Speech. *The
Information Society*, 12:365-386.

Phillips, David J. 1996. "Defending the Boundaries: Identifying and Countering Threats in a Usenet
Newsgroup", *The Information Society*, 12:39-62.

Poster, Mark. 1990. *The Mode of Information: Poststructuralism and Social Context* Chicago: University
of Chicago Press.

Quarterman, John S. The Matrix: Computer Networks and Conferencing Systems Worldwide, Bedford
Massachusetts: Digital Press, 1990.

Raymond, Eric (editor). 1993. "The On-Line Hacker Jargon File" (ver. 3.0.0). Electronic document.
(FTP: rtfm.mit.edu). Also published as "The New Hacker's Dictionary" (2nd ed.). Cambridge,
MA: MIT Press.

Reid, Elizabeth. 1996. "Informed Consent in the Study of On-Line Communities: A Reflection on the Effects of Computer-Mediate Social Research". *The Information Society*, 12(2) 169-174.

Rice, Ronald E, "Communication Networking in Computer-Conferencing Systems: A Longitudinal Study of Group Roles and System Structure," Communication Yearbook 6, M Burgoon (ed.) Beverly Hills: Sage, 1982), pp. 925-944.

Rice, Ronald E, et al. Individual and Network Influences on the Adoption and Perceived Outcomes of Electronic Messaging, Social Networks 12 (1990) 27-55.

Rice, Ronald. 1995. "Network Analysis and Computer-Mediated Communication Systems". In *Advances in Social Network Analysis*. Stanley Wasserman, Joseph Galaskiewicz, Thousand Oaks: Sage.

Rodgers, Everett M. Progress, Problems and Prospects for Network Research: Investigating Relationships in the Age of Electronic Communication Technologies. Social Networks 9 (1987) 285-310.

Salzenberg, Chip, et al. 1992. "What is Usenet?" Electronic document. (FTP: rtfm.mit.edu).

Sacks, Harvey, Schegloff, Emanuel A., Jefferson, Gail (1974). A simplest semantics for the organization of turn-taking for conversation. Language, 50, 696-735.

Schelling, Thomas C. The Strategy of Conflict. 1960.

Shibutani, Tamotsu. Improvised news; a sociological study of rumor. Indianapolis, Bobbs-Merrill 1966

Sproul LS and Kiesler, SB. 1991. *Connections: New Ways of Working in the Networked Organization*. Boston, MA: MIT Press. 205 pp.

Terveen, Loren, Will Hill, Lynn Cherny, and Steve Whittaker. "Quantifying Online Conversation" Unpublished manuscript. 1997

Turkle, Sherry. 1996. *Life on the Screen* New York: Simon and Schuster.

Waskul, Dennis and Mark Douglass, "Considering the Electronic Participant: Some Polemical Observations on the Ethics of On-Line Research", in The Information Society, 12, 129-139, 1996.

Wasserman, Stanley and Katherine Faust. 1994. *Social Network Analysis: Methods and Applications* Cambridge University Press.

Weber, R. 1984. Computer-aided Content Analysis. *Qualitative Sociology*, 7(1-2), 126-147.

Wellman, Barry, et al. Computer Networks as Social Networks: Collaborative Work, Telework, and Virtual

Community. Annual Review of Sociology 1996 Vol. 22, p 213-238

Whittaker, Steve, "Talking to Strangers: An evaluation of the factors affecting electronic collaboration",

ATT Labs, 1996.

Whyte William H. 1971. *City: Rediscovering the Center*

---

[2] Conferencing systems are distinguished from alternative network interaction media like email, discussion lists, chat, muds, and graphical worlds, by their collections of messages created by a group of people that are organized around topics, "threaded" into chains of responses and replies, and exchanged asynchronously. Other conferencing systems include PicoSpan (used on the WELL), Caucus, and a variety of BBS systems. The major commercial services, AOL, CompuServe, and MSN also offer discussion groups.

[3] Broadly, it can be argued that all kinds of digital artifacts, like bank balances, credit reports, and architectural designs are social. But not all cyberspaces are social in the sense of primarily providing a venue for interaction between people.

[4]The system is publicly accessible at http://netscan.sscnet.ucla.edu/. David Faraldo and Alex Brown were instrumental in implementing Netscan. This study was conducted using data collected from the UCLA campus news server (news.ucla.edu). News servers collect and store messages in Usenet newsgroups and send specific messages in response to requests made by news clients. News is exchanged via the Network News Transport Protocol (NNTP) and is often managed by a server known as Internet Network News (INN). See ftp://ftp.internic.net/rfc/rfc1036.txt for definition of message headers, and ftp://ftp.internic.net/rfc/rfc977.txt for definition of NNTP.

[5] Usenet interactions are embedded in a larger set of communications ties. Most Usenet participants have access to alternative channels of communication (email, IRC, MUDs, WWW, telephone, face-to-face interaction, etc.) none of which is visible in the data Netscan uses. In many cases, interactions that start in Usenet newsgroups are continued in more private email discussions.

[6] Through most of the history of the Usenet some of its content has been archived, but much of the Usenet is routinely erased after a period of a few days or weeks. As a result, there are no clear long-term measures of its growth, content and dynamics. Although some of the data can be reassembled from scattered pieces, a complete picture of the history of Usenet may be difficult to reconstruct.

[7] This number is actually a count of the distinct email addresses used to post messages to the Usenet.

[8] Because of crossposting, many of these messages were duplicated in more than one group. If messages are counted each time they occur in any newsgroup, more than 19 million messages were distributed. This means that about 40% of the Usenet is composed of redundant information. This value may be even higher as 9% of all Usenet messages are quoted text (segments of prior messages) reproduced in responses (Brewer and Johnson 1996).

[9] Although commercial providers like AOL and MSN claimed 9 million and 2 million users respectively in 1997, it is likely that these figures are significantly inflated and do not reflect the real number of users who participate in these system's conferencing areas. Furthermore, these systems are far more geographically bounded (serving mostly North America and Europe) than the Usenet.

10 Zero messages originated from the following domains Antarctica, Bangladesh, Burundi, Buthan, Bouvet Island, Belize, Congo, Guadeloupe (Fr.), Equatorial Guinea, Guam (US), Guinea Bissau, Guyana, Comoros, St.Kitts Nevis Anguilla, Korea (North), Liberia, Lesotho, Northern Mariana Isl., Martinique (Fr.), Malawi, Norfolk Island, Nauru, St. Pierre & Miquelon, Palau, Rwanda, Solomon Islands, Svalbard & Jan Mayen Is, Chad, Tadjikistan, Tokelau, Tunisia, Tanzania, Virgin Islands (British), Wallis & Futuna Islands.

11 The value for the United States is a total of the com, net, edu, org, us, gov, mil, and usa domains. This may not be entirely accurate, many com domains may actually be accessed from outside of the United States.

[12] For ten years Brian Reid of the DEC West Coast Research Lab ran a program that generated reports that came to be known as "Arbitrons". Through the cooperation of a large group of system administrators, Reid

received information on the number of newsgroups, the number of postings in each group, the number of bytes in those posts, and the number of "readers" for each group. It is the measure of readership that was most often contested. Reid's method involved a sample that may not have been representative and used multipliers to estimate the total size of the Usenet population that may have been dramatically off. Still, even if the magnitudes of Reid's data are mistaken, the process he used to measure the Usenet remained relatively constant throughout his study making the relative measures useful. For a discussion of the many potential flaws in these measurements see: http://www.tlsoft.com/arbitron/statistical_error.html

[13] Network measurement is a fairly new and somewhat inexact art in which significant differences of opinion and methodology continue to exist. Estimates of the number of people online, for example, vary from tens of millions to over one hundred million people. While the exact figures are important, the general trend is indisputable; people are adopting computer-networked media at an impressive rate. (See: RFC1296, www.mids.com, www.nw.com and ftp://ftp.isoc.org/isoc/charts/hosts3.ppt). Sampling methods and survey research may offer some insight into readership patterns, a possible future direction for further research.

[14] Crossposting creates a non-directional tie between two or more newsgroups. The data included in each message's header does not identify from which group a post originates. As a result, all connections between newsgroups are bi-directional. No data is available to determine if a group exports more messages than it imports.

[15] Netscan collects the From, Date, Newsgroups, References, Message ID, Hosts, and Subject elements from the header of every message in every group and creates a database from the information. This database was used to generate the measures presented here. In this study I ignore the body of the messages exchanged, and focus on the structure of social interaction, not its content. Analyzing message content in aggregate can be a difficult task because of the vast storage requirements imposed by the heavy message traffic in the Usenet. However, message content can be investigated using widely available computer textual analysis methods that can be quite sophisticated (Weber 1984; Damashek 1995). These techniques

represent a direction for future work. The results of computer textual analysis methods can not replace ethnographic study of social cyberspaces but can be a powerful complement to such research.

[16] Usenet newsgroups can be "moderated" which allows only messages specifically approved by the newsgroup's owner to be posted to the newsgroup. Only about 1544 (10%) of newsgroups are moderated. Many Usenet users are aware that the moderation feature can be fairly easily circumvented.

[17] There are reception filters, known as kill files, which allow a particular poster's messages to be screened out of a newsgroup for a particular reader. The posts remain available for others who have not filtered out that poster.

[18] Network interaction systems like the Usenet continue to grow in use even in the face of newer and flashier network media where others have become extinct. The World Wide Web has all but killed the older Gopher protocol. But the Usenet, email, IRC and MUDs seem to have a future as integrated components of the WWW.

[19] During the history of the Usenet, waves of new users have flooded in repeatedly, first as universities opened access to the Internet and thus the Usenet to student populations, and again as commercial online services and Internet Service Providers have made network access a consumer commodity. Each wave has challenged the existing structure and traditions of the Usenet and resulted in the fashioning of new ones.

[20] Attacks and chronic problems include spam, cancelbots, topic and newsgroup name drift. Spam is the term used for messages that are mass mailed or posted throughout the Usenet without regard to the topics of the groups. Cancelbots have been developed as a response to this kind of shotgun littering. They are automated methods for deleting offensive messages. However, the solution may be more of a problem than its inspiration since they introduce the possibility of censorship. Name and topic drift occurs when the content of a discussion shifts but its subject title does not. This erodes the coordination system that keeps the millions of messages sensibly organized.

[21] One application of this research may be to provide better clues as to where these islands are located. Related research and applications identify the web sites most frequently mentioned in a newsgroup (Hill

and Terveen 1996). Other applications focus on ways to collect and apply the reactions of prior readers of online materials to allow groups to review and recommend material to one another automatically (Brewer and Johnson 1996).

[22] With the exception of newsgroups in the "alt" hierarchy, new newsgroups are created through a fairly elaborate electoral process. In exchange for following this process ratified newsgroups gain wider distribution. While no Usenet site is required to accept or pass along all newsgroups, by informal agreement many sites carry any newsgroup that passes the electoral hurdle. In contrast, newsgroups in the "alt" hierarchy can be created at a moment's notice by anyone who desires to. The trade-off is that many Usenet sites refuse to carry "alt" newsgroups entirely or only carry select newsgroups. Many "alt" newsgroups must develop a moderately strong following before they will be widely distributed, posing a chicken-and-egg like start-up problem. Still, many "alt" newsgroups succeed and are widely available.

[23] The problem is complicated by the fact that some news servers will list newsgroups that they do not actively carry. A possible solution to this problem is to connect Netscan to multiple servers or to run multiple copies of Netscan at sites around the world. These copies could exchange information with one another to generate as complete a picture of the entire Usenet as possible.

[24] This level of coverage seems to be fairly equal with a range of commercial providers of Usenet feeds for consumer use. Few servers attempt to collect a feed for every possible group, waiting for a request for a particular group instead.

[25] This figure may be significantly inflated. The count of empty groups reflects the message holdings of the single newserver used in this study. In some cases newsgroups are created globally but contain messages that are only distributed regionally, making the message traffic within them invisible from other parts of the world. Measures of non-zero newsgroups are likely to be more reliable.

[26]Many commercial Internet providers automatically direct their customers to news.newusers.questions for support, funneling thousands of people into the Usenet and this newsgroup every month.

[27] Since messages are contributed from all over the world, each time stamp has been adjusted to GMT to provide a uniform basis for comparison.

[28] Newsgroups are also connected to other groups when they share the same poster or reader regardless of whether they have been crossposted together.

[29] Network maps are created in conjunction with a social network analysis tool such as Krackplot (http://www.contrib.andrew.cmu.edu/~krack/).

[30] One indication of a group's level of organization is the existence of documents that document some of the group's goals and history. For example, in many Usenet newsgroups a Frequently Asked Questions (FAQ) contains lists of answers to the most common questions and issues raised in a particular newsgroup. These files can save the regular participants of a newsgroup the irritation of having the same topics and questions come up continuously. They also provide a valuable resource for anyone just starting to explore the topic of the newsgroup. There are 1,623 (14%) newsgroups with a FAQ listed in the main repository for FAQs, the RTFM site maintained by MIT ("RTFM" is the canonical answer to common, trivial, or annoying questions on the Internet and means roughly "Read the Manual". The site is located at ftp://rtfm.mit.edu/pub/usenet/). The remaining newsgroups may be well organized and useful, but the presence of a FAQ is an important sign of the maturation of a newsgroup.

[31] Whenever people interact with other people or organizations through any computer system they may leave a durable mark. Phone calls, banking and credit transactions, library loans, video rentals, airline reservations, magazine subscriptions, utility bills and other forms of electronic records are routinely cataloged and analyzed for a variety of purposes.

[32] I follow King's lead by avoiding all identification of individuals in this paper.

[33] Messages that contain an optional X-No-Archive Header are not archived by the Deja News service, allowing anyone who is well informed enough to opt out of the archival service. The Netscan analysis tool also honors this convention.