The changing relationship between information technology and society

By Marti A. Hearst
University of California, Berkeley
hearst@sims.berkeley.edu

Society and information technology are rapidly co-evolving, and often in surprising ways. In this installment of "Trends and Controversies," we hear three different views on how society and networked information technology are changing one another.

Becoming socialized means learning what kinds of behavior are appropriate in a given social situation. The increasing trend of digitizing and storing our social and intellectual interactions opens the door to new ways of gathering and synthesizing information that was previously disconnected. In the first essay, Jonathan Grudin—a leading thinker in the field of computer-supported cooperative work—points out that, like a naive child, information technology often ignores important contextual cues, and tactlessly places people into potentially embarrassing situations. He suggests that as we continue to allow computation into the more personal and sensitive aspects of our lives, we must consider how to make information technology more sophisticated about social expectations, and become more sophisticated ourselves in understanding the nature of computer-mediated services.

In the second essay, I discuss a related issue—how newly internetworked information technology allows people acting in their own self-interest to indirectly affect the experiences of other people. It is to be expected that people will try to trick or deceive systems that support intrinsically social activities, such as running auctions. What is surprising here is that technologies that do not obviously have a social aspect, such as information-retrieval ranking algorithms, are nevertheless being manipulated in unexpected ways once they "go social."

In our third essay, Barry Wellman—a sociologist and an expert in social network theory—explains how the structure of social networks affects the ways we live and work. He describes the move away from a hierarchical society into a society in which boundaries are more permeable and people are members of many loosely knit groups. He introduces the notion of glocalization: simultaneously being intensely global and intensely local. Wellman describes how computer-mediated communication is contributing to this glocalization transition in social habits and infrastructure. As networked information technology continues to provide us with new views of ourselves, we hope that these essay will help designers of information technology better understand the broader impact of the work they do.

Has the ice man arrived? Tact on the Internet

Jonathan Grudin, UC Irvine and Microsoft Research

Several years ago at Bellcore, researchers thought it would be great to access newsgroup contributions by people they admired. They wrote a program to archive and search newsgroups. They tested it by entering the names of a few colleagues. "We soon found," one recounted, "that we were discovering things about our friends that we didn't want to know."

The Internet has created a new focus of computation: computer-mediated communication and interaction. Most of what is communicated is received indirectly. On the Web, above all else we see what people produce and make available; also, we read what people say and how others respond, receive indications of what people have done or are doing, and so on. The Internet's greatness resides in this extremely efficient spread of information. It is efficient, but it is not discrete, not tactful. Even when communicating directly on the Internet, we often neglect tact for brusqueness or flaming. Indirect communication and awareness, the focus of this essay, is unsoftened by the technology.

A word to the wise

Human communication is marked by tact. Knowing when and how to be tactful requires knowledge of the communication context, which is often lost or altered in computer-mediated interaction. Newsgroup messages are written in a context that appears to participants to be "chatting in a room," an ephemeral conversation among a group of like-minded people. But of course what is said can later be read outside that context, by anyone, anytime, anywhere. It can even end up being read in court.

Is anything wrong with openness? Is tact necessary? Well, yes, it is. The candor of children, who don't fully understand a conversation's social context, can be refreshing in small doses, but we all learn that tact is essential in most communication. We constantly observe social conventions, avoid social taboos, minimize needless embarrassment, and allow people to preserve the gentle myths that make life more pleasant. Eugene O'Neill's play *The Ice Man Cometh* outlines a series of calamities that occur when his characters are briefly forced to abandon these myths.

Consider another example, in which technology removed an illusion of fairness. A programming class instructor proposed that students submit homework solutions and receive the graded corrections via e-mail. The students produced a counter-proposal: After grading an exercise, the instructor posts all of the graded solutions for everyone to see! In this way, the students can discover what had been tried, what worked and what didn't, and which solution is more elegant. They can learn from each other.

It sounds great. But, those who have graded papers probably recall that after working through the entire set, you might regrade the first few, because it took a while to work out how many points to subtract for this or that kind of error. Grading is not perfectly consistent. In this class, the grading is visible to everyone. The instructor works harder than usual to be consistent, but students still detect inconsistencies, complain, and might conclude that a previously admired instructor is careless or unfair. The instructor works harder than usual to be consistent, but ends up disappointing the students. The students' illusion, their belief in the consistency of grading, is undermined by the technology. It is tempting to welcome a dose of reality, but in these examples, no one is happy about the outcome.

Another example: Compliment a conference organizer on the smoothness of the event and you might be told, "If you could see the chaos and near-catastrophe behind the scenes...." Now, technology can reveal what had been "behind the scenes." In the Web's early days, I participated in two conferences in which much of the activity was made visible to all program committee members. For example, once reviews of submissions were written, we could read them online and begin resolving differences of opinion by e-mail, prior to the program committee meeting. Very efficient, but problems arose.

Administrative errors in handling the database were immediately seen by everyone and led to confusion or embarrassment. Reviewers could scan the reviews and observe patterns: for example, you were invariably easy, I was invariably harsh; she was a careful reviewer, he was pretty casual about it. In addition, some reviewers felt uneasy about their reviews of a paper being read "out of context" by people who had not read the paper. Assumptions of smooth management and comparable reviewing performance were demolished. The planning of these conferences seemed chaotic to me, but one of the organizers remarked that in his experience, it was in fact unusually smooth, because the organizers knew that all slipups would be visible and thus "we felt we were on stage at all times; we had to be careful." Our difference in perception arose because the technology made visible more of the underlying reality.

The underlying reality

What is the underlying reality? Ethnographers or anthropologists have studied workplaces and repeatedly shown that behavior is far less routine than people believe. Exception-handling, corner-cutting, and problem-solving are rampant, but are smoothed over in our reports and even in our memories, whether out of tact or simply to get on with the job. People normally maintain an illusion of relative orderliness.

Technology is changing that. The more accurately and widely it disperses information about the activities of others, the more efficiently we can work, but at a price: irregularity, inconsistency, and rule breaking that were always present are now exposed and more difficult to ignore. In a well-known example, technology could detect all automobile speeding violations. If we don't use it, how do we decide when and against whom to selectively enforce the law?

A police officer might use context to guide enforcement—weather and traffic conditions, perhaps. We might tactfully overlook a colleague's occasional tardiness. But technology is poor at assessing context; it does not tactfully alter a time stamp. We once could imagine a colleague as an imposing person, who pays attention to detail, but e-mail reveals his careless spelling, his outdated Web site instantly reveals a relative lack of organization or concern for his image, and a video portal catches him picking his nose. None of this negates the huge benefits of these technologies, but it creates a challenge. Many challenges, in fact: in our computer-mediated interactions during the days and years to come, we will have to address this issue over and over, as individuals, as members of teams and organizations, and as members of society.

What to do?

How can we address technology's lack of tact, its inability to leave harmless illusions untouched?

Can we build more tact into our systems? Spelling correctors help. Perhaps the video portal, detecting a colleague changing clothes for a tennis match and having forgotten about the camera, could recognize what is happening and discretely blur the focus. Perhaps a virtual Miss Manners could proofread my e-mail, or a virtual lawyer could scan an automatically archived meeting and flag sensitive words. But realistically, these are exceedingly subtle, complex, human matters involving knowledge of an interaction's context, tacit awareness of social conventions and taboos, and appreciation of which illusions and corner cutting are harmless or even beneficial and which are problematic. It is a worthy goal, but intelligent systems will only slowly start to carry some of the weight.

Coming Next Issue

Intelligent Rooms

In our next issue, Haym Hirsh will present a discussion of intelligent rooms, with essays by

- **James Flanagan, Rutgers University**
- **Michael Mozer, University of Colorado**
- **Richard Hasha, Microsoft**
- **Michael Coen, MIT AI Lab**

Another possibility is to retreat. In some cases, we will decide the increased efficiency isn't worth it. In the examples I've cited, the newsgroup scanner was abandoned, the conferences stopped making as much information visible in subsequent years, and posting graded exercises has not become a custom. But these were intentionally extreme examples. Examples abound in the daily use of the Internet and Web, from which there will be no retreat. Our actions are becoming much more visible; the global village is arriving. And, in general, I believe there are tremendous benefits in efficiency, in the fairness that visibility promotes, and in the ability to detect significant problems and inconsistencies. We might be too worried, too cautious in embracing these new technologies.

A third approach seems inevitable: We will find new ways to work, to organize ourselves, and to understand ourselves. The solutions might not be obvious. I have frequently described the case of the programming class instructor, who works harder but has a more dissatisfied class, as an apparently insoluble dilemma. I recently presented it to Douglas Engelbart. He thought for several seconds, then said, "The class could develop a collective approach to grading assignments."

When information technology "goes social"

Marti Hearst, UC Berkeley

In everyday life we often observe the unintended consequences of the actions of individuals on society as a whole. If I intend to go to San Francisco from Marin County, I might well get in my car and drive to the Golden Gate Bridge. Although I certainly do not have the goal of slowing down someone else's trip to the city, my action might indeed contribute to this result. I can even unintentionally add hours to the travel time of thousands of fellow motorists if my car stalls on the bridge. Most people do not ever consider deliberately blocking traffic, but there are exceptions. Protestors can exploit the vulnerability of the freeway system to tie up the rush-hour commute, and youths can deliberately disrupt local traffic patterns by "cruising" suburban streets.

The rise of the Web and other networked information technologies has brought about new, sometimes surprising, ways for the actions of individuals and small groups to have impact on other people with whom they otherwise have no relationship. Many of these new opportunities are exciting and promise great benefits. For example, after I purchase a book from Amazon.com, I am shown suggestions of books bought by other people who also bought my new book. If I want to find out how to fix an electrical problem with my car, it may be the case that someone I never met has written up a solution and placed it on the Web.

However, the interconnectivity and global accessibility of the Web has also given rise to some unexpected ways in which people can take advantage of the technology at the expense of other people. Applications that heretofore would not have been assumed to have social ramifications are in fact allowing unexpected interactions among their users. This essay presents the case that information scientists need to begin thinking about design in a new way—one that incorporates the potential consequences if the output of their systems are likely to "go social." Information technology "goes social" when the exposure of its output makes a transition from individuals or small groups to large numbers of interconnected users.

Gaming Web search engines

Let's look at a few examples. The first is a field I know well—information retrieval.

The standard problem in IR is that of helping users find documents that (partially) fulfill an information need. If there were only a few documents to choose from, finding the relevant ones would be a simple process of elimination. However, there are millions of valuable documents as well as myriad documents of questionable general worth (for those who think the Web contains mainly junk, the Library of Congress alone catalogs over 17 million books, and a trend toward moving materials online will ensure large amounts of high-quality online material). Given many equally valid pieces of information coexisting simultaneously, the problem becomes that of pushing aside those that are not relevant, or pulling out the few that are relevant to the current need. Thus it is not so much a problem of finding a needle in a haystack, as finding a needle in a "needlestack."

Information technology "goes social" when the exposure of its output makes a transition from individuals or small groups to large numbers of interconnected users

IR is different than retrieval from a standard database-management system. In DBMSs, all information is entered in a precisely controlled format, and for a given query there is one and only one correct answer. By contrast, IR systems must make do with only an approximation to an accurate query, ranking documents according to an estimate of relevance. This fuzzy behavior is an unfortunate consequence of the fact that automated understanding of natural language is still a distant dream.

Instead of understanding the text, an IR algorithm takes as input a representation of the user's information need, usually expressed as words, and matches this representation against the words in the document collection. In practice, if the query contains a relatively large number of words (say, a paragraph's worth), then documents that also contain a large proportion of the query words will tend to be relevant to the query. This works because there tends to be overlap in the words used to express similar concepts. For example, the sentence "The Mars probe Pathfinder is NASA's main planetary explorer" will tend to share words with a newspaper account of the same topic. However, this strategy is not infallible; if an inappropriate subset of query words overlaps, nonrelevant documents may be retrieved. For example, an article containing the sentence "A vandal easily mars the paint job of the Pathfinder, the Explorer, and the Probe" shares four terms with the previous sentence, although their meanings are quite different.

Additionally, the short length (1-2 words) of queries submitted to search engines could cause IR systems to retrieve documents unrelated to the user's information need. For example, a user searching for articles on Theodore Roosevelt might find information about a football team located at a school named after this US president.

Thus IR systems circumvent the need for automated text understanding by capitalizing on the fact that the representation of a document's contents can be matched against the representation of the query's contents, yielding inexact but somewhat usable results. For over 30 years, IR research has focused on refining algorithms of this type. However, in the course of those 30 years, no one had the faintest glimmer of what would happen when IR technology went social.

What had never been imagined was that authors would deliberately *doctor the content* of their documents to *deceive* the ranking algorithms. Yet this is just what happened once the Web became widespread enough to be attractive to competing businesses, and once search engines began reporting that thousands of documents could be found in response to queries.

Web-page authors began gaming the search-engine algorithms using a variety of methods. One technique is to embed the contents of the wordlist of an entire dictionary in the Web page of interest. (The words are hidden using the HTML comment tag—comments are invisible to humans reading the page, but are indexed by some Web search engines. A similar effect can be achieved by formatting the text in the same color as the page background.) For the reasons I've described, the inclusion of additional words, whether or not they have anything to do with the content of the page, increases the likelihood of a match between a user’s query and a Web page.

There are also cases of authors placing words that are known to be of interest to many information seekers (“sex” or “bug-free code,” for example) into a Web page’s meta tag field, because some search engines assign high weight to meta tag content. A variation on this theme is to use a word that really is relevant to the content of the Web page, but repeat the word hundreds of times, exploiting the fact that some search engines increase a document’s ranking if a query term occurs frequently within that document. Listing the names of one’s competitors in the Web page’s comments section can also mislead a search engine; if a user searches on a competitor’s name, the search engine will retrieve one’s own Web page but no information about the competitor will be visible.

These techniques could be seen as modern-day equivalents of naming businesses in such a manner as to get them listed first in the phone book— AAA Dry Cleaners, for example. This doctoring of the content of documents might also be considered an entirely new way of using words as weapons; a new way to make words mean other than what they say; something we might call subliminal authoring.

Search-engine administrators quickly catch on to these techniques. Ranking algorithms can be adjusted to ignore long lists of repeated words, and some search engines do not index comments or meta tags because of the potential for abuse. This can quickly devolve into a series of moves and counter-moves. For example, users can submit Web-page URLs to search engines to get the pages reindexed and thus have the index reflect changes more rapidly. Some Web-page doctorers (incorrrectly) assumed that multiple submissions of a page would cause its ranking to increase, and so tried submitting their pages thousands of times over. Search-engine administrators noticed this behavior and started taking punitive action against repeat resubmitters. In response, some people have considered repetitively resubmitting the Web pages of their competitors in the hopes of getting these pages eliminated from the search engine indexes.[1]

Of course, search-engine providers aren’t all innocent in this. It is claimed that some will rank Web pages higher than others for a fee. This kind of behavior is also something that simply would not have been thought of in the earlier, pre-social days of information retrieval.

System design for social interactions

The lower levels of networking software allow computers to send and receive data from one another. The difficulties with such software reside in the design of systems that work accurately, reliably, and efficiently.

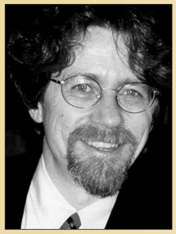
However, it has become apparent that the difficulties in the design of systems that support interaction among groups of people or on behalf of people lie not so much in the creation of efficient, reliable algorithms. Instead, these systems must be designed to take into account fuzzier concerns relating to the social practices, assumptions, and behaviors of people. Computer-supported cooperative work (CSCW) researchers have shown that groupware applications such as shared calendars and meeting tools must be sensitive to the various conflicting goals of the group participants. For example, administrative assistants, engineers, and managers disagree on what the important features of a calendar/scheduling system are.[2]

Information systems that take actions on behalf of human users must take into account how users might try to manipulate the system. Designers of auction or voting systems must consider how users might try to deceive the system by voting multiple times or preventing others from voting. Designers of agents that negotiate prices for goods must consider the potential for bait-and-switch pricing tactics, pricing collusion between competitors, and general fraudulent business practices. Because these systems perform actions traditionally done by people interacting with one another, it is perhaps unsurprising (in retrospect) that social considerations must be taken into account to make these systems succeed.

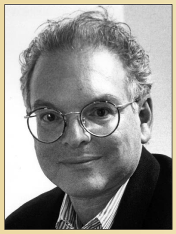
The new phenomenon we observe here is that even systems whose underlying goal is not that of supporting social interactions are nevertheless being used in this manner. We might need to accede that when information technology goes social, information-system developers must learn to adopt defensive strategies, just as neophyte drivers have to learn about defensive driving. Defensive driving is not necessary if there are no other drivers on the road; similarly we do not need this type of defensive strategy with information technologies unless they are networked together.

What’s in a domain name?

Let’s now consider another example. A Web-page server’s “real” network address



Jonathan Grudin is a professor in the Information and Computer Science Department of the University of California, Irvine, and a member of the Collaboration and Education Group at Microsoft Research. His research interests include human-computer interaction and computer-support cooperative work. He earned a BA in mathematics and physics from Reed College, an MS in mathematics from Purdue University, and a PhD in cognitive psychology from the University of California, San Diego. He is Editor-In-Chief of *ACM Transactions on Computer-Human Interaction* and was co-chair of the CSCW’98 Conference on Computer Supported Cooperative Work. Contact him at Microsoft Research, One Microsoft Way, Redmond WA 98052-6399; grudin@ics.uci.edu; www.ics.uci.edu/~grudin.



Barry Wellman is a professor of sociology at the University of Toronto, the founder of the International Network for Social Network Analysis, the chair of the Community section of the American Sociological Association, and the virtual community focus area advisor of ACM Siggroup. He attended the Bronx High School of Science in the days of rotary calculators and learned how to keypunch and wire a counter-sorter while getting his PhD from Harvard. He has coedited *Social Structures: A Network Approach* (2nd ed., JAI Press, London, 1997) and has edited *Networks in the Global Village* (Westview Press, Boulder, Colo., 1999). He is spending January through May as a visiting professor at the School of Information Management and Systems, University of California, Berkeley. Contact him at the Centre for Urban & Community Studies, Univ. of Toronto, 455 Spadina Ave., Toronto, M5S 2G8 Canada; wellman@chass.utoronto.ca; http://www.chass.utoronto.ca/~wellman.

Marti Hearst is an assistant professor in the School of Information Management & Systems at the University of California, Berkeley. She is also coeditor of “Trends & Controversies.” Her research interests focus on user interfaces and robust language analysis for building information-access systems, and on furthering the understanding of how people use and understand such systems. She received her BA, MS, and PhD in computer science from the University of California, Berkeley. Contact her at the School of Information Management & Systems, South Hall, Rm. 102, Univ. of California, Berkeley, CA 94720-4600; hearst@sims.berkeley.edu; http://www.sims.berkeley.edu/~hearst/.

is a represented as a string of digits separated by periods. These serve as identifiers to allow computers on the network to distinguish one from another.

However, Web servers also support URLs that contain domain names, which act as mnemonic pseudonyms for the numeric IDs. Usually, a domain name reflects the name of the institution to which it belongs. For example, *www.berkeley.edu* refers to the UC Berkeley home page, and *www.whitehouse.gov* refers to the US White House's home page.

An entirely unexpected and opportunistic exploitation of these naming conventions has arisen, relying on the fact that people tend to make spelling errors. Web sites have been created whose domain names have no resemblance to the content they contain, or whose domain names are common misspellings of the names of popular sites. For example, *www.whitehouse.**com*** contains pornographic material; conversely, *www.playby.com* consists solely of advertisements for technical products.

Names are not particularly important when a computer is communicating with another computer. Within computer systems, ID strings serve simply to distinguish one entity from another and do not have intrinsic meaning. However, once exposed to and used by people, the symbols take on meaning. People will interpret and interact with the identifiers in ways impossible to imagine a computer doing. Most likely the creation of mendacious domain names would not have been thought of, much less considered important, until large numbers of people became interconnected, using not only the same technology but also viewing the same information.

This situation stems in part from the rather egalitarian manner in which domain names were originally assigned. In fact, domain names were allocated in a manner similar to how the Department of Motor Vehicles assigns vanity license plate names. Pretty much anyone can have pretty much any license plate as long as it isn't already taken by someone else and fits within the prescribed length limitations and uses the standard alphanumeric characters. License-plane names are also subject to certain restrictions about what constitutes good taste, and it has long been a game of the public versus the DMV to try to fool the censors into accepting license plates with questionable interpretations.

The difference between URLs and license plates, of course, is that only a few people can see a license plate at any one time, and they are not particularly useful for business on a large scale. Also, a car cannot be instantly retrieved just by invoking the name on it's license plate.

Hypertext

I am a member of an interdisciplinary program whose faculty include computer scientists, law professors, economists, and other social scientists, and whose mission is to integrate the social and the technical in the study of the organization, management, and use of information.

One day in lecture last semester, I mentioned to our interdisciplinary masters students that HTML and the Web ignored much of what had been learned about hypertext in the preceding decade, including such things as link types and bidirectional links. One student asked what would happen if the Web allowed bidirectional links. I did what all smart professors do when posed with a difficult question in class: instead of answering, I made it into a homework assignment question.

What had never been imagined was that authors would deliberately doctor the contents of their documents to deceive search-ranking algorithms.

I asked the students to perform a *gedanken* experiment, and discuss what would happen if the Web supported bidirectional links. They were to consider a scenario in which, if a link was made from A to B on any page, a reverse link could be forced to appear from B to A.

In my computer scientist naivete, I assumed this would be a good thing, allowing me to easily show citations at the end my text and have the citations point back to the place in the text from which they were referenced, make it easier to make tables of contents, and generally make it easier to find related information.

However, the socially savvy students' answers surprised me. Out of 19 students, only one thought bidirectional links would be an inherently good thing. Instead, they foresaw all manner of disastrous outcomes, including

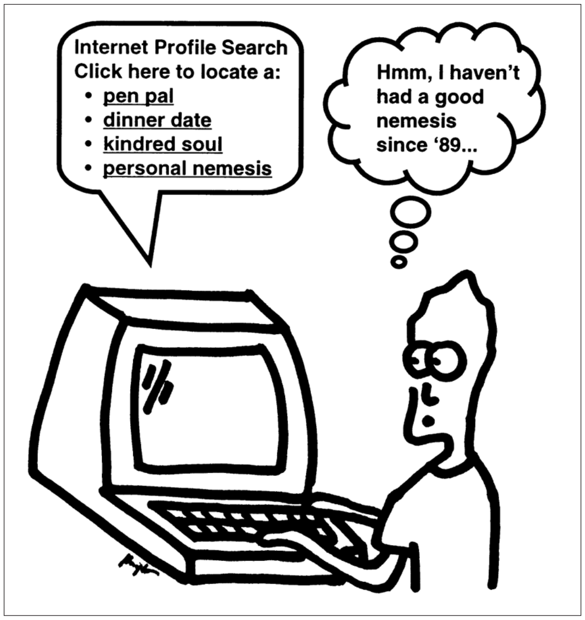
- *Link spamming:* for example, people could damage a company by flooding its home page with spurious backlinks, or people could force someone's personal home page to link back to an offensive page about themselves (such as "babes of the Web").
- *False endorsements:* people could make it look as if some entity endorsed their Web page by linking to that entity; pages could be forced to link to advertisers' pages.
- *Loss of control of information:* If bidirectional links were the only type of link available, their use could prevent the ability to hide internal information, as in the case in which a link internal to a firewall pointed to a page in the external world.

Of course, no one has suggested implementing forced bidirectional links in this way (the standard technical solution is to store all links in a separate link database, rather than place them within the page itself). On the Web, standard read/write restrictions on file systems prevent this kind of activity. However, when discussing why bidirectional links were not used in the design of HTML and HTTP, these kinds of concerns are not named. In the design notes for the WWW, Tim Berners-Lee writes:

> Should the links be monodirectional or bidirectional?
>
> If they are bidirectional, a link always exists in the reverse direction. A disadvantage of this being enforced is that it might constrain the author of a hypertext—he might want to constrain the reader. However, an advantage is that often, when a link is made between two nodes, it is made in one direction in the mind of its author, but another reader may be more interested in the reverse link. Put another way, bidirectional linking allows the system to deduce the inverse relationship, that if A includes B, for example, that B is part of A. This effectively adds information for free. ...[3]

Here, Berners-Lee expresses concern about a lack of control by the author over the reader's experience, but none of the potentially negative social impacts considered by my students comes into account.

Before going social via the Web, most hypertext linking happened within a single document, project, or small user group. In the late 80's, before the rise of the Web, there were many competing technologies, none compatible with the others Since going social, hypertext has become useful for linking information in far-flung places, assembled by people who don't know each other or have access to one another's data. Links outside of given projects can be more useful than internal ones, because they lead to resources less likely to be known to the internal group members. However, this kind of interaction was not on the radar screen in hypertext thought and research.



For example, in the ACM Hypertext 89 proceedings,[4] the authors were generally concerned with semantics of link types, navigation paths, how not to get lost (still a big problem!), and how to author hypermedia documents. Only two papers discuss the possibility of cross-project links. The first, a systems paper by Amy Pearl describing how documents on different systems might be interlinked, simply assumed bidirectional links as the only link type. The other paper, called "Design Issues for Multi-Document Hypertexts" by Bob Glushko, shows clearly that at the time the notion of inter-document linking in real systems was a radical one.

In his closing plenary address at Hypertext 91,[5] Frank Halasz revisited the issues he had raised in his landmark 1987 paper "Reflection on NoteCards: Seven Issues for the Next Generation of Hypermedia Systems." These issues related to searching, link structure, and various computational concerns. Halasz also discussed supporting social interactions over a hypermedia network, but focused on Randy Trigg and Lucy Suchman's notion of mutual intelligibility[6] (making sure participants can understand what each person is doing) and how to write readable hypertext (which in retrospect he realized did not belong in the social category). Halasz also introduced four new issues, one of which was the need for open systems to allow cross-system linking, and another which he called the problem of very large hypertexts. The problems he foresaw in this category had to do with scaling large systems and disorientation in large information spaces. He did not mention potential social concerns.

A book called *Society of Text*[7] published in 1989 contains a collection of 23 research papers on hypertext, multimedia, and their use. However, no papers discuss the consequences of many simultaneous users, or even begin to hint at the possibility of deceitful or ill-intentioned linking. Rather, hypertext was discussed in terms of how it might bring about a new way of thinking, a way of modeling the mind in the computer, or a new way of reading. Most of the concern was about how to design hypertext layout to eliminate confusion and clutter. The social concerns pertained to how the writing profession might change and how users collaborate when authoring together.

Given that it still wasn't clear if hypertext would even be intelligible to most people, it is perhaps not surprising that researchers were not considering what would happen when millions of people were linking hundreds of millions of documents.

Ted Nelson, who coined the term "hypertext" in 1965 and who since then has been an evangelist for its execution in his vision of the Xanadu system, did worry about certain social issues, namely copyright and how to handle payments for access (this system was the subject of a critical legal analysis by Pamela Samuelson and Robert Glushko, which brought up additional social issues[8]). In the Xanadu system, authors were to pay to put their writings in the system, and readers were to pay to read these works. Readers could also add hyperlinks to improve the findability of information within the system, and would receive payment when other readers used these links. Link creators would only be compensated if their links were traversed by others, thus motivating authors to create high-quality links. However, pernicious links like those anticipated by the SIMS students were not considered, perhaps because Xanadu was to be a closed system over which its administrators could exert control.[8]

A true exception can be found in Jakob Nielsen's 1990 book *Hypertext & Hypermedia.*[9] On page 197 of this book of 201 pages, under the heading "Long Term Future: Ten to Twenty Years," he cautiously predicts large shared information spaces at universities and some companies. In this context, he points out some potential social consequences of shared information spaces.

> If thousands, or even millions of people add information to a hypertext, then it is likely that some of the links will be "perverted" and not be useful for other readers. As a simple example, think of somebody who has inserted a link from every occurrence of the term "Federal Reserve Bank" to a picture of Uncle Scrooge's money bin. ...
>
> These perverted links might have been inserted simply as jokes or by actual vandals. In any case, the "structure" of the resulting hypertext would end up being what Jef Raskin has compared to the New York City subway cars painted over by graffiti in multiple uncoordinated layers.[10]

Interestingly, three paragraphs later, he also proposes the use of popularity of following hyperlinks as a measure of the usefulness of the link, but does not consider the possible gaming effects using this technology, as I discuss next.

Collaborative ratings

Information technology going social can open up new opportunities. Many researchers and developers have noted that information technology allows for the tracking and logging of the information seeking behavior of masses of users. One oft-stated suggestion is to gather information about preferences by users' implicit choices, by keeping track of which hyperlinks are followed, which documents are read, and how long users spend reading documents. It is hypothesized that this information can be use to assess the popularity, importance, and quality of the information being accesses, and used to improve Web-site structure and search-engine ranking algorithms. Again, unanticipated behavior might undermine the integrity of these systems. If the results of these algorithms lead to commercially important consequences, such as changing a site's ranking within search results, then people will be likely to write programs to simulate users' visiting the Web pages of interest, and countermeasures will be required.

Researchers are also making use of explicit rating information, most notably in what is known as collaborative-filtering systems or recommender systems.[11] Collaborative-filtering systems are based on the commonsense notion that people value the recommendations of people whose recommendations they have agreed with in the past. When new users register with a collaborative-filtering system, they are asked to assign ratings to a set of items (such as movies, recipes, or jokes). Their opinions are then matched against those of others using the system, and similar users are identified. After this, the system can recommend additional items to the new users, based on those that have already been rated by similar uses.

Collaborative filtering is a social phenomenon. Researchers have discussed some of the social dilemmas that can work to the detriment of such systems, especially issues having to do with motivating people to be initial reviewers rather than waiting for others to create the ratings.[11]

However, as we've seen, there are less obvious kinds of interactions that can degrade the system's behavior, which arise only because large masses of people use the same system.

In a recent manuscript, Brent Chun points out the motivations people might have for deceiving the system and some ways in which they might carry out this deceit.[12] He proposes that companies whose services are being rated might attempt to affect the ratings they receive or downgrade the ratings of their competitors, specific interest groups might try to further their causes by giving negative ratings to companies or products that conflict with their beliefs, and collaborative-filtering companies themselves might try to sabotage the ratings of their competitors. Chun suggests ways people might attack the ratings databases, including conventional security threats such as breaking into the system to steal or modify the database. He goes on to discuss more ingenious means for defrauding these systems, such as rating the same item multiple times using large numbers of pseudonym identities, borrowing other users' identities, and collusion within groups of authentic users to downgrade an item's rating.

Why does this happen?

These behaviors seem to occur only when a large cross-section of society uses the same technology and information simultaneously, and when that information is of general interest. As I've noted, defensive driving is not necessary if there are no other drivers on the road.

What is interesting about the phenomena described here is that social interactions occur with technology whose use does not obviously result in such interactions. It was not obvious that the Web would lead to gaming of information-retrieval systems, nor that the domain-name facility would lead to deceptive naming practices.

Which kinds of technology are susceptible to this kind of behavior? We can draw distinctions between technologies that are self-consciously about interactions among individuals and groups, and those that ostensibly have no reason to consider collusion and arms races.

Here I will venture a classification. Three conditions must hold for this situation to arise.

- First, the system must network a large cross-section of society, the members of which have partially conflicting goals.
- Second, there must be value associated with use of the system, power, prestige, financial, and so on.
- Third, and least obvious, the technology must involve human use of information in some human-understandable form.

Ramifications for information systems design

The introduction of social forces onto the landscape of information technology brings up issues that are foreign to traditional computer-science training.

Computer scientists are taught to anticipate and handle all possible kinds of input, but not at the level of granularity necessary to address these considerations. Programmers check the data type (string, integer, object pointer) and the ranges that can be taken on by these data types. A programmer learns to test for very long strings and empty strings, and perhaps whether or not a string matches a string that has been defined internally, but does not consider inquiring into whether or not the content of the string represents something socially unacceptable, something deceitful or something fraudulent. Notions of read/write protection and computer security control who has access to machines and data, but do not attempt to control fraudulent or deceitful use of the technology.

Perhaps, however, this is not the proper role of the designer of an information technology system. After all, we don't want word processors that censor what a user is typing. It could be argued that most of the interesting behaviors discussed above arise because documents on the Web are not monitored and controlled by the social norms that are usually associated with publishing. It might also be the case that it should be left to the legal system to prevent certain forms of unfair business practices that result from networked information systems. (This has already begun to happen in some cases.[13])

The importance of the interdisciplinary field of human-computer interaction is gradually achieving increased recognition within traditional computer science. HCI advocates the design of computer systems from a human-centric viewpoint, and advises us on how to create systems that "generate positive feelings of success, competence, mastery, and clarity in the user community."[14] Clare-Marie Karat has gone so far as to proclaim a User's Bill of Rights that underscores the design goals of HCI.[15] Ben Shneiderman and Anne Rose suggest that designers of information systems create "Social Impact Statements," modeled after Environmental Impact Statements, to help ensure that the technology we create achieves its intended goals while at the same time serving human needs and protecting individual rights.[16] Their framework emphasizes the importance of defining the stakeholders of the system—not just who will use it directly, but also who will be indirectly affected by its use. Now that a wider range of information technology is going social, designers should begin to consider whether or not the stakeholders are everyone.

Acknowledgments

This essay has benefited from conversations with Hal Varian, Doug Tygar, Bob Glushko, Jonathan Grudin, Brent Chun, Jef Raskin, and the SIMS masters students. Thanks also to Don Kimber.

References

1. "Zdnet: Search Engines Battle the New Spam: Make Sure Your Meta Tags Don't Get You Disqualified," *ZD Net Internet MegaSite Magazine*, Jan. 19, 1998; http://www.zdnet.com/icom/content/anchors/199801/19/new.spam/index.html
2. J. Grudin and L. Palen, "Emerging Groupware Successes in Major Corporations: Studies of Adoption and Adaptation," *Proc. First Int'l Conf. Worldwide Computing & Its Applications (WWCA97): Emerging Technologies for Network Computing*, ACM Press, N.Y., 1997, pp. 142-183.
3. T. Berners-Lee, "Design Issues," http://www.w3.org/DesignIssues/.
4. *Proc. ACM Hypertext 89*, ACM Press, 1989.
5. F.G. Halasz, "Seven Issues: Revisited., Hypertext '91 Closing Plenary," http://www.parc.xerox.com/spl/projects/halasz-keynote/transcript.html.
6. R. Trigg, L. Suchman, and F. Halasz. "Supporting Collaboration in NoteCards," *Proc. Computer-Supported Cooperative Work Conf. (CSCW'86)*, ACM Press, 1986, pp. 147-153.
7. E. Barret, ed., *Society of Text*, MIT Press, Cambridge, Mass., 1989.
8. P. Samuelson and R.J. Glushko, "Intellectual Property Rights for Digital Library and Hypertext Publishing Systems: An Analysis of Xanadu," *Hypertext*, 1991, pp. 39-50.
9. J. Nielsen, *Hypertext & Hypermedia*, Academic Press, New York, 1990.
10. P. Resnick and H.R. Varian, "Recommender Systems—Introduction to the Special Section," *CACM*, Vol. 40, No. 3, 1997, pp. 56-58.
11. C. Avery, P. Resnick, and R. Zeckhauser, "The Market for Evaluations," to appear in the *American Economic Review*; http://www.si.umich.edu/~presnick.
12. B. Chun, "Security in Collaborative Filtering Systems," 1998, http://www.cs.berkeley.edu/~bnc
13. S.M. Abel and C.L. Ellerbach, "Trademark Issues in Cyberspace: The Brave New Frontier," Fenwick and West LLP, San Francisco, 1998; http://www.fenwick.com/pub/publications.htm.
14. B. Shneiderman, *Designing the User Interface* (3rd Ed.), Addison Wesley, Reading, Mass., 1998.
15. C.M. Karat, "Guaranteeing Rights for the User," *Comm. ACM*, Vol. 41, No. 12, Dec. 1998, pp. 29-31.
16. B. Shneiderman and B. Rose, *Social Impact Statements: Engaging Public Participation in Information Technology Design*, Tech. Report CS-TR-3537, Univ. of Maryland, College Park, Md., 1995; http://www.cs.umd.edu/TRs/authors/Ben_Shneiderman.html.

Living networked in a wired world

Barry Wellman, University of Toronto

The world is composed of networks—not groups—both computer networks and social networks. When computer networks connect people and organizations, they are the infrastructure of social networks. Just as a computer network is a set of machines connected by a set of cables (or airwaves), a social network is a set of people (or organizations or other social entities) connected by a set of socially meaningful relationships (see Figure 1). Although this might be obvious to many computer scientists, the implications of living in a networked world are nonobvious.

Computer scientists have been centrally involved in a paradigm shift, not only in the way we think about things but in the way that society is organized. I call it the shift from living in "little boxes" to living in networked societies.[1] I am going to describe its implications for how we work, commune, and keep house, using the neologism called *glocalization*. Members of little-box societies only deal with fellow members of each of the few groups to which they belong: usually our homes, neighborhoods, workgroups, and organizations. We are moving away from a group-based society to a society in which boundaries are more permeable, interactions are with diverse others, linkages switch between multiple networks, and hierarchies (when they exist) are flatter and sometimes recursive.

The little-boxes metaphor is that people are socially and cognitively encapsulated by all-confining, socially conforming groups. Most people think of the world in terms of groups, boundaries and hierarchies.[2] They see themselves as belonging to a single work group in a single organization; they live in a household in a neighborhood; they belong to a kinship group (one each for themselves and their spouses) and to voluntary organizations such as churches, bowling leagues, and the Computer Society. All of these social structures appear to be bodies with precise boundaries for inclusion (and therefore exclusion). Each has an internal organization that is often hierarchically structured: supervisors and employees, parents and children, pastors and churchgoers, the Computer Society executive and its members. In such a little-box society, we only deal with the people in each of our bounded groups when we are participating as members of that group.

We have moved from hierarchically arranged, densely knit, bounded groups to less bounded and more sparsely knit social networks. (Actually, a group is a type of social network, one that is tightly bounded and densely knit, but it is cognitively easier to compare groups with more loosely bounded and sparsely knit networks.) Empirical observation has shown this shift in many milieus. Instead of hierarchical trees, management by network has people reporting to shifting sets of supervisors, peers, and even nominal subordinates. Unless

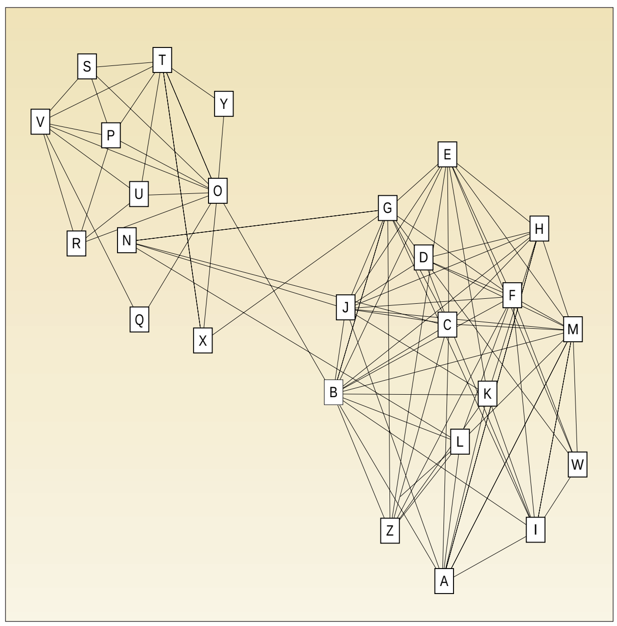


Figure 1. A social network where the labeled boxes represent individuals (people, organizations) and the lines represent relations between them (which could be love, money, or influence, for example). Note that some of the boxes (such as F and W) have two lines connecting them; they are tied by two relations. The graph shows two densely knit clusters with crosscutting ties (B and O, for example). (Courtesy of Cathleen McGrath, Jim Blythe, and David Krackhardt; *http://www.andrew.cmu.edu/user/cm3t/groups.html*.)

people are tethered to assembly lines, their work relations often spill over their work group's boundaries, and may connect them to outside organizations.[3] Organizations sometimes distrust salespeople and purchasing agents because their jobs entail more contact outside the organization than within it: they are what sociologists call marginal people who are not totally committed to one group.

Rather than belonging to a single family, people often have complex household relations, with stepchildren, ex-marital partners (and their progeny), and multiple sets of in-laws. People in the western world usually have more friends outside their neighborhood than within it: indeed, many people have more ties outside their metropolitan areas than within it.[4] People might attend several churches, and they may have to decide between going bowling or attending a Computer Society meeting. In short, while people think they are members of social groups, they more often function as operators of their social networks. (See the sidebar for pointers to more information about social network analysis.)

Although I have constructed my argument as a then/now contrast, people have always functioned in social networks to some extent. Consider the comings-and-goings in Jane Austen's preindustrial England. Whatever their "sense and sensibility," her novel's characters are forever galloping past their neighbors to visit their far-flung friends and relatives.[5] Kenneth Scherzer has shown that guests traveled considerable distances to attend New York City weddings in the mid-19th century: they provided us with empirical proof by signing the wedding registers as witnesses.[6] The telegraph enabled alert businesspeople by the middle of the 19th century to manage their affairs at a distance,[7] intensifying a tradition of spatially dispersed business empires that traders had heretofore held together through kinship loyalties and written correspondence.[8] Even before the coming of computer-mediated communication, cars, planes and phones maintained far-flung relations. For example, in the 1960s and 1970s, North Americans' important ties of sociability and support rarely were confined to their neighborhoods. Many were on the other side of the metropolitan area; some were across the continent or the ocean.[4,9]

Characteristics of computer networks as social networks

CMC—such as the Internet, newsgroups, and videoconferencing—makes it easier to be socially networked.[10]

- CMC is usually asynchronous, allowing people in different time zones or on different schedules to communicate. For example, although the computer scientists our group has studied work in the same office, their different work schedules leads them to use e-mail.[11]
- CMC is rapid, fostering a high velocity of exchanges, sometimes ill-considered.[12]
- CMC supports emotional, nuanced, and complex interactions, belying early fears that it would be useful only for simple, instrumental exchanges.
- CMC has taken on its own norms, procedures and ethos, with CMC participants showing greater creativity and emotional swings than those talking face to face. [13]
- The absence of direct feedback in most CMC encourages more extreme forms of communication. People input messages to screens that they would never say to another person palpably present in person or on the telephone.
- The ability of communications to be forwarded supports transitivity, as when messages get forwarded to friends of friends. The inclusion of headers in forwarded messages allows indirect ties to become direct relationships. This aids the exchange of information that cuts across group boundaries. Such crosscutting ties link and integrate social groups, instead of such groups being isolated in tightly bounded little boxes.
- E-mail, the only widely available form of CMC, supports easy accessibility. This has led to a leveling of perceived hierarchies, with all feeling they have access to all. E-mail is not unique in this. Telephone networks also support easy accessibility, so much so that busy and reclusive people have constructed social (secretaries) and technical (voicemail) barriers to access. CMC will probably engender the same reaction, once techno-euphoria fades, with agents both providing background detail about callers and keeping unwanted callers at a distance.
- The ease of sending messages to large numbers of recipients allows participants to remain in contact with multiple social milieus.
- E-mail is especially useful for maintaining contact with "weak ties"—persons and groups with whom one does not strong relationships of work, kinship, sociability, support, or information exchange. Because weak ties are more socially heterogeneous than strong ties, they connect people to diverse social milieus and provide a wider range of information.[14]
- CMC's accessibility, velocity and multiple-message characteristics indirectly connect the entire world in five steps or less.[15] Yet, unlike computer networks in which all nodes are ultimately connected, there is significant decoupling in social networks. Hence, information diffuses rapidly through computer-supported social networks, but neither universally nor uniformly.[16,17]

Too many studies have focused too tightly on CMC without realizing that its interactions are only part of the life that extends beyond the screens.

Although most experimental studies of CMC look only at screen-to-screen relationships, people who relate to each other online often relate to each other offline: face-to-face, by phone, or even on paper.[12] Many studies have focused on CMC without realizing that such interactions are only part of the life that extends beyond screens. We can only comprehend the role of CMC if the total tie is taken into account, and not just the on-screen relationship. For example, our group's study of a wired suburb has found that extremely fast and accessible Internet access spurs neighborly interaction as well as far-flung ties. Neighbors use the Internet to arrange get-togethers and to organize in opposition to real estate developers. This neighborly interaction is not surprising: Until wireless CMC becomes prevalent, people are largely tied to their computers at their office or home desktops. This is the phenomenon we call glocalization: the situation of being intensely global as a Net surfer while being firmly rooted to the area around the computer screen and keyboard.[18,19]

Computer-mediated interactions are socially situated as well as spatially situated. Thinking of computer networks as social networks can move the study of human computer interactions beyond looking only at the standard HCI concerns of person-screen or person-screen-screen-person interactions. Even when only two persons communicate, they are not dancing duets in isolation. Their interactions are conditioned by the availability of others to supply resources, cause problems, or enforce norms.

Moreover, unlike laboratory experiments of CMC, in real life CMC is often between people who have different social characteristics—such as gender and lifestyle—and different social positions—such as supervisors and subordinates and core and periphery. For example, our group's study of desktop videoconferencing saw supervisors initiating more contact than subordinates. Some coworkers in a separate office 100 kilometers away maintained autonomy when their videoconferencing equipment frequently "broke down."

Even when there is unfettered computer connectivity, not all persons or organizations are directly connected. A computer network is not in itself a social network: it is the technological infrastructure that enhances the ability of people and organizations to communicate for better or worse.

How does living in networks differ from living in groups?

(1) It enhances the ability to connect with a large number of social milieus, while decreasing involvement in any one milieu.
(2) It decreases the control that any one social milieu can have over us, while decreasing the commitment of any one milieu to a person's well-being.
(3) It shifts interactions from those based on characteristics people are born with—such as age, gender, race and ethnicity—to characteristics that they have adopted throughout the life course—such as lifestyles, shared norms and voluntary interests.
(4) It fosters "cross-cutting" ties that link and integrate social groups, instead of such groups being isolated in tightly bounded little boxes.
(5) It has increased choices while reducing the palpable group memberships that provide a sense of belonging.
(6) In short, it has reduced the identity and pressures of belonging to groups while increasing opportunity, contingency, globalization, and uncertainty through participation in social networks. ■

Social network analysis

To learn more about the discipline of social network analysis, see the following sources:

- S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications,* Cambridge Univ. Press, Cambridge, UK, 1993.
- J. Scott, *Social Network Analysis,* Sage, London, 1991.
- B. Wellman, "An Electronic Group is Virtually a Social Network," *Culture of the Internet,* S. Kiesler, ed., Lawrence Erlbaum, Mahwah, N.J., 1997, pp. 179–205.
- B. Wellman and S. D. Berkowitz, eds. *Social Structures: A Network Approach, 2nd ed.,* JAI Press, Greenwich, Conn., 1997.

The International Network for Social Network Analysis (INSNA) has a Website: *http://www.heinz.cmu.edu/project/INSNA/.*

References

1. M. Reynolds, *Little Boxes,* Schroeder Music, New York, 1963.
2. L. Freeman, "The Sociological Concept of Group: An Empirical Test of Two Models," *Am. J. Sociology*, Vol. 98, 1992, pp. 152-66.
3. B. Wellman et al., "Computer Networks as Social Networks: Virtual Community, Computer Supported Cooperative Work and Telework," *Ann. Rev. Sociology*, Vol. 22, 1996, pp. 213-238.
4. B. Wellman, ed. *Networks in the Global Village,*Westview Press, Bolder, Colo., 1999.
5. J. Austen, *Sense and Sensibility*, Penguin, Harmondsworth, UK, 1811 (1969).
6. K. Scherzer, *The Unbounded Community: Neighborhood Life and Social Structure in New York City, 1830-1875,* Duke Univ. Press, Durham, N.C., 1992.
7. A. Pred, *Urban Growth and the Circulation of Information: The United States System of Cities, 1790-1840,* Harvard Univ. Press, Cambridge, Mass., 1973.
8. N. Ferguson, *The World's Banker,* Weidenfeld & Nicolson, London, 1998.
9. C. Fischer, *To Dwell Among Friends*, Univ. of California Press, Berkeley, Calif., 1982.
10. L. Garton and B. Wellman, "Social Impacts of Electronic Mail in Organizations: A Review of the Research Literature," *Communication Yearbook*, Vol. 18, 1995, pp. 434-53.
11. C. Haythornthwaite and B. Wellman, "Work, Friendship and Media Use for Information Exchange in a Networked Organization," *J. Am. Soc. Information Science*, Vol. 49, 1998, pp. 1101-1114.
12. J. B. Walther, J. F. Anderson, and D. W. Park, "Interpersonal Effects in Computer-Mediated Interaction: A Meta-Analysis of Social and Antisocial Communication," *Comm. Research*, Vol. 21, 1994, pp. 460-487.
13. L. Sproull and S. Kiesler, *Connections,* MIT Press, 1991.
14. M. Granovetter, "The Strength of Weak Ties: A Network Theory Revisited," *Social Structure and Network Analysis*, P. Marsden and N. Lin, eds., Sage, Beverly Hills, Calif., 1982, pp. 105-130.
15. H. White, "Search Parameters for the Small World Problem," *Social Forces*, Vol. 49, 1970, pp. 259-64.
16. T. Valente, *Network Models of the Diffusion of Innovations,* Hampton Press, Cresskill, N.J., 1995.
17. B. Wellman and S.D. Berkowitz, eds. *Social Structures: A Network Approach,* 2nd ed., JAI Press, Greenwich, Conn., 1997.
18. K.N. Hampton and B. Wellman, "Netville: Glocalization and Civic Society," presented to *American Sociological Association*. San Francisco, 1998 (contact author for copies).
19. B. Wellman and M. Gulia, "Net Surfers Don't Ride Alone," *Networks in the Global Village*, B. Wellman, ed., Westview Press, 1999, pp. 331-366.
20. J. Scott, *Social Network Analysis*, Sage, 1991.
21. S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*. Cambridge Univ. Press, Cambridge, UK, 1993.
22. B. Wellman, "An Electronic Group is Virtually a Social Network," *Culture of the Internet,* S. Kiesler, ed., Lawrence Erlbaum, Mahwah, N.J., 1997, pp. 179-205.