Modeling and Supporting Communities on the Internet and World Wide Web

Brian R. Gaines and Mildred L. G. Shaw
Knowledge Science Institute
University of Calgary
Alberta, Canada T2N 1N4
{gaines, mildred}@cpsc.ucalgary.ca

Abstract

Arguably the most important societal impact of the Internet is its support for special-interest communities. Communication through email and list servers and shared information archives through web browsers and servers support the discourse and knowledge processes of on-line communities in such a novel and effective way as to be revolutionary. Tools are being built that add to the effectiveness of such communities by organizing their knowledge products, supporting awareness of new material, attracting relevant members, and so on. User models need to be developed on which to base the development of effective tools, but these models need to move beyond individual cognitive processes to provide models of user communities and of the relationships between individual and community processes. This article describes research on community processes, the underlying cognitive, cultural, social and media theories, empirical modeling of community processes, and the use of the results to characterize user needs, the dynamic development of user models and support through tools.

Key words: Internet community modeling and support, socio-cognitive theories and tools

1 Introduction

Our research focus has long been the study and support of human-computer interaction, but it has gradually shifted from a focus on individuals and their cognitive processes to groups and their cognitive processes. In the past decade the growth of the Internet has widened the focus to encompass groups that are so large and diffuse as to best be conceptualized as *virtual communities* (Rheingold, 1993). Some groups that we have been supporting and studying have been professional communities defined only by their common interests for whom we, or others, provide services. Other have been major international projects with better defined membership related to an agenda of specific tasks. What we have come to realize is that the effective support of such communities depends on modeling not only the cognitive processes of individuals, but also those of the community as a whole. Many of the tools we have developed support community processes, and the most significant and challenging of these are those that model these processes dynamically so that the community can reflect upon its own operations.

A typical community that is intermediate between the well-defined team considered in computer-supported cooperative work (CSCW) and the diffuse professional community using list and web servers to coordinate itself is the GNOSIS intelligent manufacturing systems (IMS) consortium that we have studied as a 'society of research agents' (Gaines and Norrie, 1997). The GNOSIS consortium was formed in 1992 to develop a long-term IMS research program concerned with the systematization of knowledge for design and manufacturing. It comprised 31 organizations in 14 countries, involved over 100 researchers, and made heavy use of the Internet to support its research activities.

Several issues became apparent in support of GNOSIS and other communities that led us to add additional features to the Internet services:-
- Members did not find it easy to keep track of changes to the web site and maintaining a manual 'what's new' list was labor-intensive. We developed the CHRONO tool for reverse

chronological indexing of web sites to generate automatic lists of new accessions (Chen and Gaines, 1996).

- The web site needed to be mapped in terms of the conceptual structures of the community. For GNOSIS we experimented with the use of knowledge acquisition tools to extract these structures from the project planning documents (Gaines and Shaw, 1994) and generated concept maps that provided hypertext links to material on the web (Gaines and Shaw, 1995).

- Newcomers found it difficult to understand the state of the project (while GNOSIS had about 60 members at any one time, turnover in the members was such that some 100 people were involved in the first year). We added hypermail indexing of the mail archives to give access to past discourse, and textual indexing of the documents to make them accessible through keyword search.

- In the more diffuse communities newcomers from different disciplines found the terminology and concepts being used difficult to understand. We moved our conceptual modeling tools to the web to allow newcomers to compare their conceptual models with those of existing group members (Shaw and Gaines, 1995).

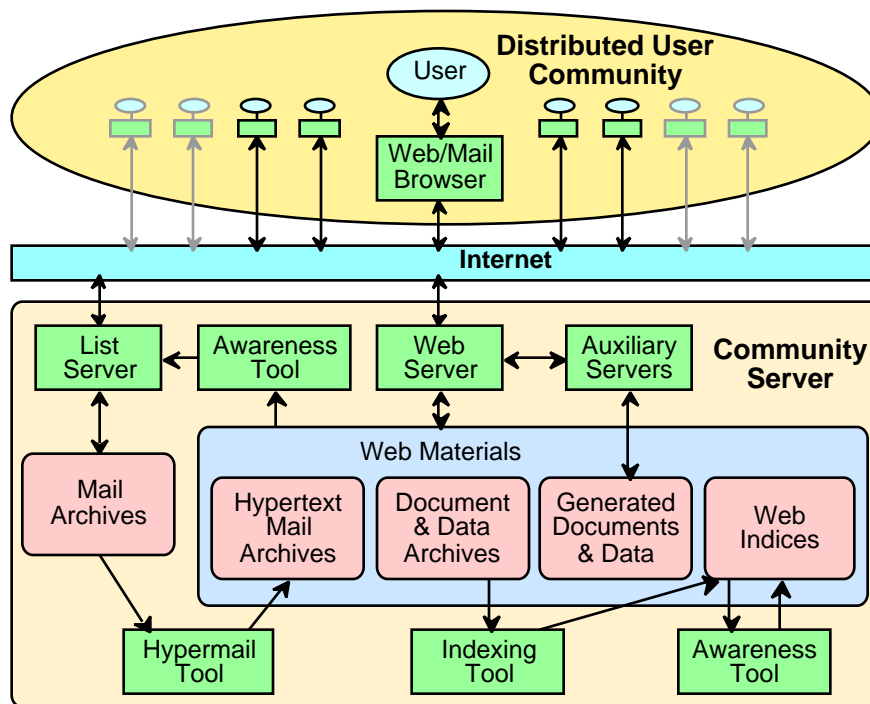Figure 1 shows the architecture of the community servers we have been using:

Figure 1 Community server for distributed Internet community

- Members of the distributed user community communicate with one another and community servers through the Internet using standard web browsers that support email and web protocols.
- A list server supports discourse within the community as a whole and sub-communities.
- A web server supports access to documents and data, including members posting documents to the server.
- A hypermail tool is used to put the list server archives on the web, indexed by date, sender and subject.
- An indexing tool is used to index the web documents by content and provide keyword search.
- Chronological awareness tools are used to monitor changes to documents and data, reporting them through the web or by selective email.

- Auxiliary servers provide web-based conceptual modeling tools.

Integrated services such as those shown in Figure 1 are becomingly increasingly used to support communities on the Internet, e.g. the GMD *BSCW* system (Bentley, Appelt, Busbach, Hinrichs, Kerr, Sikkel, Trevor and Woetzel, 1997). The remainder of this article describes approaches to modeling the user community in order to support it through such tools more effectively and to develop further tools based on dynamically generated models of the community processes.

2 Characterizing a Community in Terms of Awareness

The starting point in modeling Internet community processes is to distinguish them from individual human-computer interaction issues and from the structured workflow interactions supported in computer-supported cooperative work. What characterizes the more diffuse communities using list servers to discuss their interests and to engage in joint projects that are usually not highly structured?

In a previous article we have distinguished *teams*, *special-interest communities* and *the Internet world at large* in terms of members' *awareness* of other members (Gaines, Chen and Shaw, 1997). Does a member know others *extensionally* by who they are, or *intensionally* by their type of interest. Each resource provider in a team has an *extensional awareness* of their actual resource users, and each resource user has an *extensional awareness* of the resource and who will provide it. In a special interest community resource providers usually do not have such extensional awareness of the resource users, and, if they do, can be regarded as forming teams operating within the community. Instead, resource providers usually have an *intensional awareness* of the resource users in terms of their characteristics as *types* of user within the community. Resource users in a special interest community may have an extensional awareness of particular resources or resource providers, or an intensional awareness of the types of resource provider likely to provide the resources they require.

This model in terms of members' awareness differentiates teams from communities and draws attention to the need to model and support awareness. *Situation awareness* has proved a powerful concept in modeling the human factors of teams: "The critical thing about doing shared tasks is to keep everyone informed about the complete state of things" (Norman, 1993). User modeling in synchronous groupware may be based on the notion of workspace awareness (Gutwin and Greenberg, 1998). The primary dimensions of awareness in the more diffuse structures of Internet communities may be elicited by using the usual "wh-" interrogative pronouns that characterize the location of agent activity in physico-social space:-

	"Wh-"	Awareness	Question	Awareness Support	Tool
Product	**What**	Knowledge	What products has the community produced?	Information retrieval, Memetic tracking	Indexing
	How	Conceptual	How does the community discuss a topic?	Conceptual comparison	WebGrid
Location	**When**	Chronological	When did a significant event occur in this community?	Event tracking	CHRONO
	Where	Life-world	Where is locus of discourse currently?	Community tracking	CliqueMap
Agents	**Who**	Organizational	Who plays what role in this community?	Interaction process analysis	SYMLOG
	Why	Intentional	Why is the community undertaking this activity?	Thread tracking	ThreadMap

Figure 2 Analysis of awareness and its support in Internet communities

The following subsections discuss modeling and supporting awareness in these six categories.

2.1 Knowledge Awareness

The web materials that accrete through the processes of a community constitute its *knowledge product*, its contribution to *world 3* (Popper, 1968). The discourse of communities that do not maintain email archives and web sites have a very ephemeral existence, and most scholarly and task-oriented communities on the Internet maintain and value their knowledge products. They provide a record of the *memes* (Dawkins, 1982) that underlie the culture of the community, its *cultural software* (Balkin, 1998) existing independently of the minds that developed, and are developed by, those memes.

Many tools have been developed to index materials on the entire web or particular document collections (Marchionini, 1995), and techniques have been developed to model individual users and expedite their access to web materials (Maglio and Barrett, 1997). The techniques range from information retrieval based on content to the maintenance of selective collections of links by users representing different perspectives on community interests. In particular, a community may maintain documents answering frequently asked questions (FAQs) to help newcomers and discouraging the use of the list for elementary queries from new members. *Answer Garden* (Ackerman and Malone, 1990) provides a structured dynamic FAQ where questions not already answered are sent to an appropriate expert whose answers are posted back to the FAQ, and has been developed to operate as an *organizational memory* supporting collaborative help through the web (Ackerman and McDonald, 1998).

The GNOSIS community had a well-defined mission and we used this to index its knowledge products through layered *concept maps* in the *Mediator* system (Gaines, Norrie and Lapsley, 1995). We also experimented with generating such concept maps automatically through the textual analysis of documents describing the project through analysis of the co-occurrence of words in sentences, a technique commonly used in information retrieval systems (Callon, Law and Rip, 1986). Figure 3 shows a concept map generated from a document that played a major role in the design of the GNOSIS research program (Tomiyama, 1992), and is treated as a set of entities which are sentences whose features are the words they contain. Rules are derived using empirical induction in which the premise is that if one word occurs in a sentence then the conclusion is that another will occur. The graph shows the links from premises to conclusions derived in this way (Gaines and Shaw, 1994).
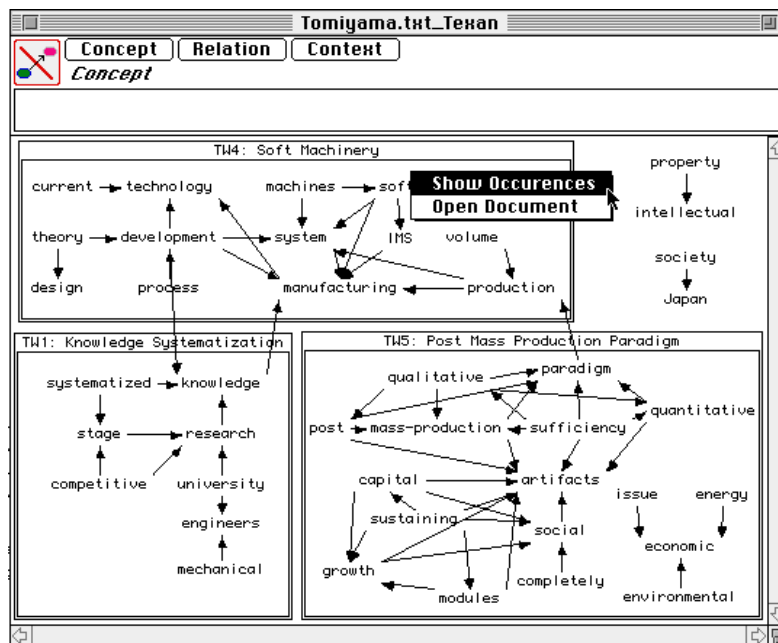

Figure 3 Concept map derived by text analysis from paper on objectives of IMS program
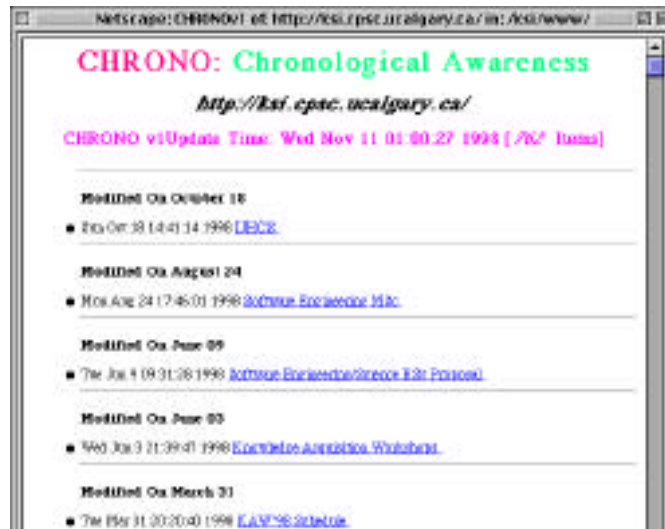
We have used such maps on the web as links to project material (Gaines and Shaw, 1995), and have developed web-based versions of the concept mapping tools that allow the maps to be edited through the web (Kremer and Gaines, 1996).

2.2 Conceptual Awareness

Communities develop conventions in the use of language that make it difficult for those outside the community to understand discourse within it. The *language game* played by a community defines the meaning of the terms being used (Wittgenstein, 1953), and any glossary is only a snapshot of a dynamic process. In scholarly communities colloquial words are often used as *aide memoires* for technical terms which are intended to evoke a highly specific context for the discourse (Roberts and Good, 1993), and members who do not know the technical term will be misled if they read it colloquially. We have developed a tool, WebGrid (Shaw and Gaines, 1995), that allows new members to check their usage of terms against those of experts in the community. Figure 4 shows a map of GNOSIS constructs and projects. WebGrid is described in detail in a companion article (Shaw and Gaines, 1999).

Figure 4 WebGrid map of GNOSIS constructs and projects

2.3 Chronological Awareness

Awareness that interesting events are occurring is important to effective participation in a community: has discussion of a new topic commenced; have new documents been posted? The appropriate support for chronological awareness depends on the expected rate of change of the area of interest. If changes are very infrequent then automatically generating email draws attention to changes without requiring user action, e.g.URL-Minder (NetMind, 1995) sends email when the content at specified URLs change. If changes are frequent then the automatic generation of a 'what's new' page is more appropriate, e.g., as shown in Figure 5, CHRONO (Chen and Gaines, 1996) automatically maintains a reverse chronological index of specified directories of a web site.

Figure 5 CHRONO automatic "what's new" chronological awareness support

2.4 Life-World Awareness

The "where" question for a virtual community needs careful consideration. There has been a debate about whether the term 'community' is appropriate for a social system that has no well-defined physical boundaries (Jones, 1997). This was an issue in sociology long before virtuality with some definitions of community involving a geographic area and others only social interaction supporting a common life based on a unity of belief and work (Hillery, 1955). From the "common life" perspective one may answer the "where" question in terms of the positioning of a community's *life-world* (Schutz and Luckmann, 1973) in cyberspace.

One can model the sub-communities within an Internet community using standard social network analysis tools to determine strongly-connected cliques in the graph of email interactions (Garton, Haythornthwaite and Wellman, 1997). However, this does not determine the location of the community within global cyberspace, i.e. how this virtual community's life-world relate to those of other virtual communities. The global location can be mapped by comparing the sets of members of various lists and regarding individuals having joint membership of two lists as providing a *weak tie* (Granovetter, 1973) that support informations flows between 'neighboring' communities. It is important to work with complete membership data rather than that derived from the discourse since 'lurkers' on one list that do not contribute to it can be monitoring it in order to post relevant information from it to another list.

2.5 Organizational Awareness

Our characterization of communities in terms of forms of mutual awareness can be developed further in terms of social and organizational psychology, in particular role theory as defined by the following five assertions (Biddle, 1979):-
1 Role theorists assert that "some" behaviors are patterned and are characteristic of persons within contexts (i.e. form *roles*).
2 Roles are often associated with sets of persons who share a common identity (i.e. who constitute *social positions*).
3 Persons are often aware of roles, and to some extent roles are governed by the fact of their awareness (i.e. by *expectations*).
4 Roles persist, in part, because of their consequences (*functions*) and because they often embedded within larger social systems.
5 Persons must be taught roles (i.e. must be *socialized*) and may find either joy or sorrow in the performances thereof.

6

In these terms, the distinction above between CSCW teams and Internet communities may be restated as that of teams tending to have well-defined prescribed roles whereas communities tend to have *emergent* roles. The lack of organization chart for the fluid structure of an Internet community makes it difficult for a newcomer to understand the discourse. Who 'owns' the community and its agenda; who are the leaders; who are the administrators; who has authority over legitimate issues; who are project leaders for particular tasks; who has a role of critic; who of expert; who of facilitator; and so on. In most Internet communities these roles exist but they, and those who occupy them, are generally not prescribed and change with the evolution of the community. Internet communities have emergent communication networks (Monge and Eisenberg, 1987) and can best be modeled through *structuration theory* (Giddens, 1986) that emphasizes the reflective equilibrium between roles constraining behavior and being created by it.

The fundamental dynamics of organizations are based on *power* and *trust*, where power is the potential to influence the behavior of others and trust is confidence in one's expectations of others (Luhmann, 1979). In teams power is initially prescribed through assigned roles whereas in communities power tends to derive from increasing trust in a member's willingness and capability to fill an emerging role. In the early development of media richness theory it was assumed that email was too impoverished to support the human interactions leading to trust, and it has been suggested that trust cannot develop in a virtual community (Handy, 1995). However, empirical studies have shown that email can provide a rich medium for human discourse (El-Shinnawy and Marcus, 1997), and that trust develops effectively in virtual teams and communities (Jarvenpaa and Leidner, 1998).

Interaction process analysis (Bales, 1950) provides a theoretical framework for the analysis of email discourse as a social network of emergent roles. Its practical application in the form of SYMLOG (Bales and Cohen, 1979) has been used to profile team activities through a time series analysis of the power and affect dimensions of the roles involved (Losada and Markovitch, 1990). ListA (Chen, 1997) is a tool supporting this analysis of list server discourse, and Figure 6 shows a SYMLOG field diagram generated by ListA of mail on an Internet list. The highly positive/dominant leadership role being played by the individual in the upper right quadrant is that of the list founder promoting the activities on the list.
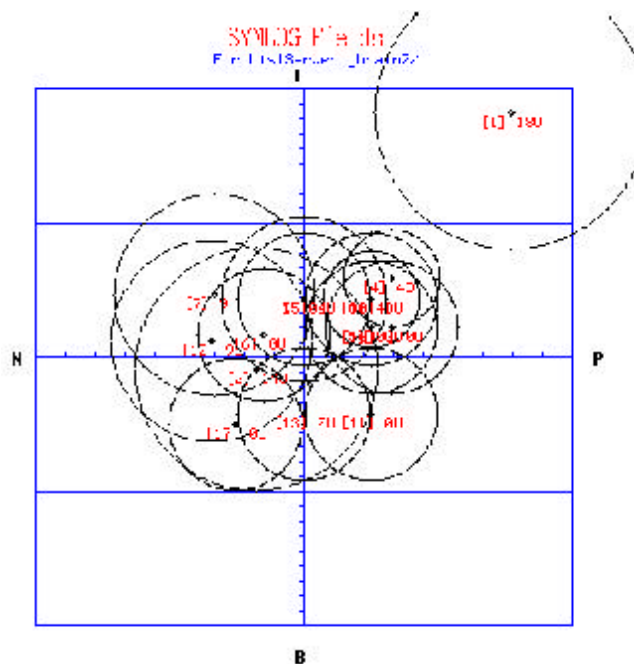

Figure 6 SYMLOG field diagram for discourse on a list server

A combined analysis of the social network and the affective dimensions enables the power structure of a group to be modeled, and those playing particular roles to be identified. Our current analysis of email lists involves manual encoding of the SYMLOG dimensions supported by computer tools and hence cannot be used for automatic modeling. We are experimenting with text analysis to determine if the major dimensions can be inferred from the affective loadings of terms in messages.

2.6 Intentional Awareness

List members tend to initiate discussions relating to particular tasks or topics and the basic intent is to promote discourse relating to a particular issue. The "Subject:" line of the message is generally used to indicate the topic, and grouping mail with the same subject line provides an elementary way of supporting intentional awareness. Some topics form recurrent memes in the culture of the community and recur without subject-line linking to earlier discussion. We are experimenting with clustering email items using standard cosine measures of term vectors to compare the items, and hence supporting thread-tracking by content rather than subject line. We have also found it useful to have the system recognize the first mail from a new user and automatically send them a welcome message giving the location of the web site and FAQ.

3 Conclusions

Internet communities have been modeled in terms of their forms of situational awareness to provide foundations for the development of systems supporting community as well as individual processes. Six forms of awareness have been distinguished: knowledge; conceptual; chronological; life-world; organizational; and intentional. The theoretical foundations for each have been discussed, and various forms of support system have been illustrated.

The analysis presented may be given an integrative framework by considering the model of an Internet community to be that of its *group mind* (McDougall, 1920). Such a *collective stance* (Gaines, 1994) is justified if the community as a whole possesses competencies beyond those of its individual members, i.e. some form of *organizational knowledge* (Gaines, 1997). The group mind concept has been used to analyze the human factors of flight operations on aircraft carriers (Weick and Roberts, 1993), and has be modeled as a transactive memory system, a set of individual memory systems in combination with the communication that takes place between individuals (Wegner, 1987). This definition captures the human components of an Internet community and may be extended to encompass the documentary material as a shared aid to human memory.

Finally, it is important to consider the ethical dimensions of the systems we have described. Members of Internet communities may not wish their processes to be modeled, and what is intended as a support tool may be seen as an unwanted intrusion (King, 1996). It is important to secure informed consent to the analyses being presented from the support systems.

Acknowledgements

References

Ackerman, M.S. and Malone, T.W. (1990). Answer Garden: A tool for growing an organizational memory. **Proceedings of the ACM Conference on Office Information Systems**. pp.31-39. New York, ACM.

Ackerman, M.S. and McDonald, D.W. (1998). Answer Garden 2: Merging organizational memory with collaborative help. **Proceedings of the Seventh Conference on Computer-Supported Cooperative Work**. pp.to appear. New York, ACM.

Bales, R.F. (1950). **Interaction Process Analysis**. Chicago, University of Chicago Press.

Bales, R.F. and Cohen, S.P. (1979). **SYMLOG: A system for the multiple level observation of groups**. New York, Free Press.

Balkin, J. M. (1998). **Cultural software : a theory of ideology**. New Haven, Conn., Yale University Press.

Bentley, R., Appelt, W., Busbach, U., Hinrichs, E., Kerr, D., Sikkel, K., Trevor, J. and Woetzel, G. (1997). Basic support for cooperative work on the World Wide Web. **International Journal of Human-Computer Studies 46**(6) 827-846.

Biddle, B.J. (1979). **Role Theory: Expectations, Identities and Behaviors**. New York, Academic Press.

Callon, M., Law, J. and Rip, A., Ed. (1986). **Mapping the Dynamics of Science and Technology**. Basingstoke, UK, MacMillan.

Chen, L.L.-J. (1997). Modeling the Internet as Cyberorganism: a Living Systems Framework and Investigative Methodologies for Virtual Cooperative Interaction. PhD (available at http://www.ucfv.bc.ca/cis/chenl/). University of Calgary.

Chen, L.L.-J. and Gaines, B.R. (1996). Methodological issues in studying and supporting awareness on the World Wide Web. Maurer, H., Ed. **Proceedings of WebNet96**. pp.95-102. Charlottesville, VA, Association for the Advancement of Computing in Education.

Dawkins, R. (1982). **The Extended Phenotype**. Oxford, Oxford University Press.

El-Shinnawy, M. and Marcus, M.L. (1997). The poverty of rich mediatheory: explaining people's choice of electronic mail vs. voice mail. **International Journal of Human-Computer Studies 46**(4) 443-467.

Gaines, B.R. (1994). The collective stance in modeling expertise in individuals and organizations. **International Journal of Expert Systems 7**(1) 21-51.

Gaines, B.R. (1997). Knowledge management in societies of intelligent adaptive agents. **Journal for Intelligent Information Systems 9**(3) 277-298.

Gaines, B.R., Chen, L.L.-J. and Shaw, M.L.G. (1997). Modeling the human factors of scholarly communities supported through the Internet and World Wide Web. **Journal American Society Information Science 48**(11) 987-1003.

Gaines, B.R. and Norrie, D.H. (1997). Coordinating societies of research agents—IMS experience. **Integrated Computer Aided Engineering 4**(3) 179-190.

Gaines, B.R., Norrie, D.H. and Lapsley, A.Z. (1995). Mediator: an Intelligent Information System Supporting the Virtual Manufacturing Enterprise. **Proceedings of 1995 IEEE International Conference on Systems, Man and Cybernetics**. pp.964-969. New York, IEEE.

Gaines, B.R. and Shaw, M.L.G. (1994). Using knowledge acquisition and representation tools to support scientific communities. **AAAI'94: Proceedings of the Twelfth National Conference on Artificial Intelligence**. pp.707-714. Menlo Park, California, AAAI Press/MIT Press.

Gaines, B.R. and Shaw, M.L.G. (1995). Concept maps as hypermedia components. **International Journal Human-Computer Studies 43**(3) 323-361.

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

Giddens, A. (1986). **The Constitution of Society : Outline of the Theory of Structuration**. California, University of California Press.

Granovetter, M.S. (1973). The strength of weak ties. **American Journal of Sociology 83** 1444-1465.

Gutwin, C. and Greenberg, S. (1998). Design for individuals, design for groups: tradeoffs between power and workspace awareness. **Proceedings of CSCW'98**. pp.to appear. New York, ACM.

Handy, C. (1995). Trust and the virtual organization. **Harvard Business Review 73**(3) 40-50.

Hillery, G.A. (1955). Definitions of community: areas of agreement. **Rural Sociology 20**(2) 111-123.

Jarvenpaa, S.L. and Leidner, D.E. (1998). Communication and trust in global virtual teams. **Journal of Computer-Mediated Communication 3**(4) http://www.ascusc.org/jcmc/vol3/issue4/jarvenpaa.html.

Jones, Q. (1997). Virtual-communities, virtual settlements & cyber-archeology: a theoretical outline. **Journal of Computer-Mediated Communication 3**(3) http://www.ascusc.org/jcmc/vol3/issue3/jones.html.

King, S.A. (1996). Researching Internet communities: proposed ethical guidelines for the reporting of results. **Information Society 12** 119-127.

Kremer, R.A. and Gaines, B.R. (1996). Embedded interactive concept maps in web documents. Maurer, H., Ed. **Proceedings of WebNet96**. pp.273-280. Charlottesville, VA, Association for the Advancement of Computing in Education.

Losada, M. and Markovitch, S. (1990). GroupAnalyzer: a system for dynamic analysis of group interaction. **Proceedings of the 23rd Annual Hawaii International Conference on System Science**. pp.101-110. New York, IEEE Computer Society Press.

Luhmann, N. (1979). **Trust and Power**. Chichester, UK, Wiley.

Maglio, P.P. and Barrett, R. (1997). How to build modeling agents to support web searchers. Jameson, A., Paris, C. and Tasso, C., Ed. **User Modeling: Proceedings of the Sixth International Conference, UM97**. pp.5-16. New York, Springer.

Marchionini, Gary (1995). **Information seeking in electronic environments**. Cambridge, Cambridge University Press.

McDougall, William (1920). **The group mind; a sketch of the principles of collective psychology, with some attempt to apply them to the interpretation of national life and character**. New York, Putnam.

Monge, P.R. and Eisenberg, E.M. (1987). Emergent communication networks. Jablin, F.M., Putnam, L.L., Roberts, K.H. and Porter, L.W., Ed. **Handbook of Organizational Communication**. pp.304-342. Newbury Park, CA, Sage.

NetMind (1995). The URL-Minder: Your Own Personal Web Robot. NetMind. http://www.netmind.com/URL-minder/URL-minder.html.

Norman, D. A. (1993). **Things That Make Us Smart: Defending Human Attributes in the Age of the Machine**. Reading, MA, Addison-Wesley.

Popper, K.R. (1968). Epistemology without a knowing subject. Rootselaar, B. Van, Ed. **Logic, Methodology and Philosophy of Science III**. pp.333-373. Amsterdam, North-Holland.

Rheingold, H. (1993). **The Virtual Community**. Reading, Massachusetts, Addison-Wesley.

Roberts, R.H. and Good, J.M.M., Ed. (1993). **The Recovery of Rhetoric: Persuasive Discourse and Disciplinarity in the Human Sciences**. Charlottesville, University of Virginia.

Schutz, A. and Luckmann, T. (1973). **The Structures of the Life-World**. London, Heinemann.

Shaw, M.L.G. and Gaines, B.R. (1995). Comparing constructions through the web. Schnase, J.L. and Cunnius, E.L., Ed. **Proceedings of CSCL95: Computer Support for Collaborative Learning**. pp.300-307. Mahwah, New Jersey, Lawrence Erlbaum.

Shaw, M.L.G. and Gaines, B.R. (1999). Supporting modeling of the social practices of other users in Internet communities. **User Modeling: Proceedings of the Seventh International Conference, UM99**. pp.submitted. New York, Springer.

Tomiyama, T. (1992). The technical concept of IMS. RACE Discussion Paper, No. RA-DP2, Research into Artifacts, Center for Engineering, The University of Tokyo.

Wegner, D.M. (1987). Transactive memory: a comtemporary analysis of the group mind. Mullen, Brian and Goethals, George R., Ed. **Theories of group behavior**. pp.185-208. New York, Springer.

Weick, K.E. and Roberts, K.H. (1993). Collective mind in organizations: heedful interrelating on flight decks. **Administrative Science Quarterly 38** 357-381.

Wittgenstein, L. (1953). **Philosophical Investigations**. Oxford, Blackwell.