THE UNIVERSITY OF CALGARY

# Modeling the Internet as CyberOrganism: a Living Systems Framework and Investigative Methodologies for Virtual Cooperative Interaction

by

Lee Li-Jen Chen

A DISSERTATION

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF COMPUTER SCIENCE

CALGARY, ALBERTA

September, 1997

**THE UNIVERSITY OF CALGARY**

**FACULTY OF GRADUATE STUDIES**

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a dissertation entitled "Modeling the Internet as CyberOrganism: a Living Systems Framework and Investigative Methodologies for Virtual Cooperative Interaction" submitted by Lee Li-Jen Chen in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Supervisor, Dr. Brian Gaines, Department of Computer Science

Dr. Larry Katz, Faculty of Kinesiology

Dr. Mildred Shaw, Department of Computer Science

Dr. Douglas Brent, Faculty of General Studies

External Reader, Dr. Thomas Calvert, Simon Fraser University

Date

# Abstract

The Internet has become a major vehicle for people to engage *virtual cooperative interaction* in which loosely associated individuals interact through a complex social network to mutual benefits. It has given new prominence to human discourse as a continuing source of knowledge. With the growth of usage of listservers and the World Wide Web, it is important to model and support the processes by which knowledge is acquired and disseminated through the Internet (i.e., *the net*). The emerging *cyberorganism* consisted of distributed intelligent agents, that is *the Internet community at large*, provides a 'cybernetic living expert system' with a scope and scale well beyond that yet conceivable with computer-based systems alone. This dissertation develops a *living systems* conceptual framework for modeling socio-technical processes on the net; describes various forms of support mechanisms on the net and categorizes them in terms of the model; applies the conceptual model to generate techniques and methodologies for investigating communication, social and knowledge processes within the Internet community; and integrates those processes in the cyberorganism framework.

This dissertation is divided into two parts. The first part introduces and develops the cyberorganism framework for modeling virtual cooperative interaction. The socio-technical origins and evolution of the net are investigated and the nature of the net as a living system is explored and examined. The author then analyzes system imperatives, structures, and processes involved in the cyberorganism. In the second part, the author investigates the conceptual framework's utility and application, and outlines techniques and methodologies for investigation of virtual cooperative interaction. Demonstrations of the utility of the cyberorganism framework correspond to the three system levels in the cyberorganism: *team*, *special interest community*, and *the Internet community at large.*

# Acknowledgments

The research reported in this dissertation has benefited from the support of many people. Thanks to my supervisor, Dr. Brian Gaines, for his encouragement, guidance, and inspiration throughout the development of this work. His constant support and interest have made this work possible. Thanks are also due to my supervisory committee members, Dr. Mildred Shaw and Dr. Larry Katz, who have given timely advice during the research development. Thanks to my defense committee members, Dr. Douglas Brent and Dr. Thomas Calvert, for taking interest in my work. Many thanks to people at the Knowledge Science Institute and Department of Computer Science at The University of Calgary for giving valuable discussion and enlightenment throughout the years. They include: Roberto Flores-Mendez, Dr. Rob Kremer, Wayne Robbins, Zhonge Xiao, Dr. Dickson Lukose, Mark Roseman, Linda Tauscher, Husam Kinawi, Fabian Gomes, Steve Franks, Carl Gutwin and Ted O'Grady. Special thanks to Dr. Douglas Norrie and Dr. Saul Greenberg for serving on my candidacy exam committee, and for their friendly advice and critique. Also, I like to thank Veronica Everton-Williams for her kind assistance in editing the dissertation manuscript. Finally, thanks to my mentors at the University of North Texas, Dr. Kathleen Swigger and Dr. Chao-Chih Yang, who first encouraged me in the areas of human-computer interaction and advanced information systems.

I could not have undertaken my graduate studies and completed this dissertation without the loving support from my family. First, I would like to express my deep gratitude to my grandma Ar-Sir and granddad Mu-Zi, for their caring affection. Great thanks to my mom and dad for their love, encouragement and direction. Thanks to my brother, Ted, for his tremendous support and discussion about scientific research. Thanks are also due to my aunts and uncles for their endearment and inspiration over the years. They include: Helen, Mei-Ying, Yu-Ming, Pochezang, Justin, Pi-Zhao, Luke, Shin-Jer, Lin-Show, Lee, Yi-Zi, and Hung-Chi. Also thanks to my godmother Bao-Yui and godfather Jin-Chang as positive role models in life. Finally, many thanks to my cousins for their friendly support; particularly to: Daphne, Carol, and  Judy for their spirited e-mail correspondence.

Dedicated to my parents, Eunice and Chang-Keng Chen

# Table of Contents

# List of Tables

# List of Figures

# CHAPTER 1

# Introduction

The overall aim of the research reported in this dissertation is to develop a model of the socio-technical dynamics of the Internet in supporting cooperative interactions and to formulate methodologies and techniques for investigating the model.

This dissertation is about modeling the Internet as a living system. A biosocial metaphor for network computing, patterned on the natural evolution of biological organisms is used through out the current work. It presents a theoretical framework for the investigation of human-computer interaction in networked computing environments. In particular, it focuses on the nature of *human-computer symbiosis* and *virtual cooperative interaction*.

Conceptually the Internet is a lot of things. It is a network of computers; more precisely, it is not just a network, but a network of networks—*inter-networks*—and hence its name. It is also software, computers, and other technologies; the net is more important in *what it enables* than *what it is* (Weintraut, 1997). It is a catalyst of change; a new medium for communication, social, and knowledge processes; and, an emergent, complex system.

Over the years the Internet (the net) has become a major resource for people to engage in *virtual cooperative interaction* in which loosely associated individuals interact through a complex social network to mutual benefits. This has given new prominence to human discourse as a continuing source of knowledge. With the growing usage of listservers and the World Wide Web, it is important to model and support the processes by which knowledge is acquired and disseminated through the net. One objective here is to provide a coherent *systems* perspective of computer-mediated interactions involving individuals, groups, and communities.

The main characteristics of systems thinking emerged simultaneously in several disciplines during the first half of the century, especially during the 1920s. Systems thinking was pioneered by biologists, who emphasized the view of living organisms as an integrated whole. It is further enriched by Gestalt psychology and the new science of ecology (Capra, 1996). Some of the central ideas of systems perspective are: connectedness, relationships, context. According to the systems view, the essential properties of an organism, or living system, are properties of the whole, which are not specific to the parts. They arise from the interactions and relationships among parts. These properties are destroyed when the system is dissected, either physically or theoretically, into isolated elements. Although we can discern individual parts in any system, these parts are not isolated, and the nature of the whole is always different from the mere sum of its parts.

Another key criterion of systems thinking is the ability to shift one's attention back and forth between systems. Throughout the living world we find systems nesting within other systems, and by applying the same concepts to different levels—for example, the concept of growth to an organism, a city, or an economy—we can often gain important insights. On the other hand, we also have to recognize that, in general, different system levels represent levels of different complexity. At each level, the observed phenomena exhibit properties of a particular level are called "*emergent*" properties, since they emerge at that particular level (Capra, 1996).

The *systems* framework characterizes the Internet as an *open, dynamic system of inter-connected machines, users, and resources*. Here, various *constructs* involved in network computing, human cognitive processes, social systems, and information retrieval/ transmission are identified and explored for their inter-relationships. As a result, the studies of the virtual cooperative interaction on the Internet draw upon concepts, theories, and research findings from several other disciplines—notably, psychology, sociology, sociobiology, and population ecology.

In this dissertation, I will employ Miller's (1978) *living systems theory* as the theoretical foundation. The theory treats individuals, groups, organizations, and communities as levels of living systems possessing many characteristics in common—namely, the essential characteristics of life itself. Within this theoretical framework various aspects of human-computer interaction (HCI) and computer-supportive cooperative work (CSCW) that were usually treated as separate and distinct, using terminology that makes it difficult to see connections between them, become part of a whole (Tracy, 1989). For example, *awareness support* becomes a basic link between coordination, information resource, discourse, motivation, reinforcement and all other issues involving virtual cooperative interaction. Together, those topics add up to different perspectives on a single entity, a living system that is the Internet.

## *1.1 Growth and Development*

The growing availability of collaborative systems and services on the Internet have expedited innovative knowledge creation/dissemination processes. These advanced information infrastructures include: digital journals, electronic libraries (Gaines, 1993a), resource discovery environments (Bowman, Danzing, Manber & Schwartz, 1994), co-authoring systems (Baecker, Nastos, Posner & Mawby, 1993) and virtual scientific communities (Schatz, 1991). A major motivation behind the current research is to investigate the nature of distributed cooperative interaction among networked collaborators who use the Internet as an integral part of their working environments.

In recent years, the number of computers connected through the Internet has grown from some 28 thousand at the beginning of 1989, to over 16 million at the beginning of 1997. Figure 1 shows data plotted from the Internet Domain Surveys undertaken by Network Wizards using a sampling methodology involving checking 1% of machines (Lottor, 1997). The 1997 figures may be put into perspective by noting that it constitutes one machine on the net for every 500 people on the planet. The growth rate has been consistently some 100% a year so that, if this were sustained, within seven years there would be one Internet computer for each person. The size and growth rate of the net have already made it a substantial medium for communication. Universities and other research organizations were the primary source of the initial growth, net access became routinely available to scholars in the late 1980s, and to the general public in North America and parts of Europe in the mid 1990s (Gaines, Chen & Shaw, 1997).


**Figure 1 Growth of Hosts on the Internet 1989-1997**

A fast evolving segment of the Internet, the web was originally conceived and developed at CERN for the purpose of assisting and facilitating collaborative interactions among high energy physicists, working at various institutions in different countries, to conduct joint research projects (Berners-Lee, Cailliau, Luotonen, Nielsen & Secret, 1994). Its conceptual ancestry can be traced to *Memex*, the archetype hypertext system for scientists (Bush, 1945). Since 1993, it has diffused at a phenomenal rate from its origin and gradually has subsumed many popular Internet communication services like USENET newsgroup, electronic mail. The original charter of the web can be summarized in the following quote:

> *The World Wide Web (W3) was developed to be a pool of human knowledge, which would allow collaborators in remote sites to share their ideas and all aspects of a common project* (Berners-Lee et al., 1994).

The popularity of the web can be attributed to its emergent growth property: the ability for a new state of being to emerge naturally from a synergy among existing systems (Kauffman, 1993; Kauffman, 1995). The web is structured such that if it was used independently for two projects, and later relationships are found between the projects, then no major or centralized changes have to be made, but the information can be linked to represent the new state of knowledge. This property of emergent growth has allowed the web to expand rapidly from its origins at CERN across the Internet irrespective of national or disciplinary boundaries. Hence the dynamics of the web are based on three fundamental notions: (i) computer-supported cooperative work; (ii) hypermedia; and (iii) trans-boundary, emergent growth (Berners-Lee et al., 1994).

The phenomenal growth of the Internet and the World Wide Web presents a stimulating avenue for research in advanced distributed information systems, especially for research into how social and psychological processes operate and are structured in networked environments of various sizes and configurations (e.g., local research networks, global scholarly communities). For instance, recently we have seen emergent behaviours and interactions on the net and psycho-social studies about them, such as: interactivity in computer-mediated communications (Rafaeli, 1988; Rafaeli & Sudweeks, 1997); flaming behaviours in newsgroups (Mabry, 1997); theatrical performance on Internet-Relay-Chat (Danet, Ruedenberg & Rosenbaum-Tamari, 1997); rhetorical dimensions and control structures in USENET newsgroups (Jones, 1997; Smith, McLaughlin & Osborne, 1997); mirrored social constructs in Multi-User-Dimension (MacKinnon, 1997); and dynamics of virtual cooperative interaction (Chen & Gaines, 1997c). Those, and many other engaging issues, are becoming important as virtual groups, and communities are rapidly forming on the Internet (Sudweeks, McLaughlin & Rafaeli, 1997).

### *1.2 Virtual Cooperative Interaction*

The growth of the web, while creating a rich new resource, also creates problems of information overload. The management of the diffuse communities collaborating through the web raises human factor issues going beyond those of the coordination of smaller, goal-directed groups with well-defined roles and tasks. What are the responsibilities of information providers in supporting users of whom they are unaware, and who may be using the information in very different ways from those originally envisioned? The web supports the collaborative activities of small work groups, but it also supports those of well-defined scholarly sub-disciplines, and those of the much less defined community at large. To study and support collaborative activities on the net, we need a conceptual framework that identifies the major distinctions among "*virtual groups*" of widely differing sizes and structures, and among the various roles that originators, recipients, and intermediaries can play (Chen & Gaines, 1997a).

The fundamental nature of social behaviours on the Internet can be characterized as **virtual cooperative interaction**. The word "virtual" has two senses here: first, it denotes the notion of *virtual space*, i.e., the cooperative interaction occurs in a non-physical space which allows participants to be situated in geographically separate locations; second, it denotes that the *intention* to engage in cooperative interaction itself may not necessarily pre-exist or be conscious. Traditional notion of groupware focuses on the first sense (tele-presence in virtual space); however, there is a need to extend the notion of cooperative interaction to encompass the latter sense of virtual cooperative interaction (Chen & Gaines, 1996a).

Quite often contribution and exchange of information resources on the net/web involve cooperative interaction without pre-planned coordination. In fact, participants on the web may have no intention to cooperate in the first place. Usually, a resource provider and a resource user are unaware of each other's existence until their first interaction. Nevertheless, the interactive process between them is still loosely cooperative in nature. It differs from the traditional team-oriented cooperation where group tasks, goals, and purposes are usually well-defined.

For example, a person who is interested in learning about the Java programming language may discover someone's personal web site full of useful Java information resources and hyperlinks to other Java related sites. The resource provider in this case is only vaguely aware that some visitors have checked out her web site, and she is satisfied that her web pages seem to attract people with similar interests in Java. Once in a while, a grateful visitor (i.e., a happy resource user) may send her a thank-you note via e-mail or/and ask questions about Java. But most of the time she is unaware of who has actually utilized her information resource.

This is an unusual form of cooperation where a resource provider might never know the identity of her resource users, but nevertheless, still continues to contribute anyway. Typically, on the web, the only feedback she may receive might be the frequency of accesses to her information resources either through log-files or counter mechanisms. What does she gain in return in such a seemingly one-way cooperative interaction? Is it simply an expression of altruism? What are some possible motivations for her to contribute to the web? In general, how would one ensure the continual contribution of an information provider? (Chen & Gaines, 1996a; 1997b; 1997c).

One of the major problems of collaboration on the web is that of *coordination signals*. In other words, how can we maintain *awareness* between remote research partners when changes occur in one location that affect activities in another. Such *situational awareness* (Norman, 1993) is an important issue for supporting task-oriented collaborative projects of research groups or organizations. At the other end of the spectrum, the issue of locating where a specific information resource is on the web, i.e., *resource awareness*, is essential for supporting the research community at large. *Group awareness* (Smith, 1994) is essential in order to provide smooth coordination among members in a collaborative project team. By extending the notion of group awareness to community awareness, the Internet, as a global collective system, has become an emergent complex system of virtual communities that transcends the traditional boundaries of both physical and social communities (Chen & Gaines, 1997a).

## *1.3 The Living Internet*

As we have observed over the years, the growth and development of the net has been very rapid with little central planning, and despite its widespread use, there is little information as yet on the social dynamics of net technologies. Many systems have been developed to cope with the information overload generated by direct access to the net (Hearst, 1997; Lynch, 1997). The wide variety of indexing and search tools now available have in common the fact that they support selective attention and awareness in the communities using the net. It would be useful to be able to analyze the design issues and principles involved in these tools in terms of the knowledge and communication processes in the virtual communities using these tools (Chen & Gaines, 1996a).

The Internet has provided major new channels for the knowledge creation and dissemination in virtual communities. Increasing international connectivity has made the net accessible to special-interest communities world-wide, and electronic mail and listservers now provide a major communication medium supporting discourse in these communities. Until recent years, limitations on the presentation quality of online file formats restricted the publication capabilities of the net to rapid dissemination of files printable in paper form. However, advances in online presentation capabilities now allow high-quality typographic documents with embedded figures and hyperlinks to be created, distributed and read online. Moreover, it has become possible to issue *active documents* containing animation, simulations, and supporting user interaction with computer services through the document interface. The major part of this functionality has become accessible through the protocols of the World Wide Web, and the web itself is seen as a precursor to an *information superhighway* subsuming all existing communications media (Gaines & Chen, 1996).

Many visions of humanity working in groups suggest the analog that people on the Internet are organized like neurons in a brain. They ask the question as to whether, when connected together appropriately (with the right rules of interconnection), the human race, with the entirety of its computers, will in fact be capable of significantly greater things than today. Certainly, a networked computer society will act as a whole, as an organism[1] (Berners-Lee, 1997). The effect of working together that some envisage is greater than that. If the whole were really to behave like an organism, then it would be beyond the wit of any individual to comprehend the state of operation of the whole. Will the organism as a whole develop its own goals and ways of achieving them? (Berners-Lee, 1997).

A living systems perspective allows us to model the web/net as an emergent, cybernetic organism, or *cyberorganism* (Chen & Gaines, 1997a). It is a global living system developing, evolving and inhabiting in an information sphere or infosphere, analogous to a biological organism living inside a biosphere. Using this *living systems* conceptual framework, we can analyze and classify the types of information systems which support communication, social and knowledge processes for virtual communities.

## *1.4 Objectives*

The growth of the Internet, while creating a rich new resource, also creates problems of information overload. The management of the diffuse communities collaborating through the net raises socio-technical issues going beyond those of the coordination of smaller, goal-directed groups with well-defined roles and tasks. Virtual cooperative interaction highlights the fact that cooperative interaction on the net involves more than tradition task-oriented, group collaboration; sometimes, the intention to engage in cooperative interaction itself may not necessarily pre-exist or be conscious, nevertheless the resulting interaction is cooperative in essence.

---

[1] It is analogous to a large-scale *society of mind* (Minsky, 1985).

The overall aim of the research reported in this dissertation is to develop a model of the socio-technical dynamics of the Internet in supporting cooperative interactions and to formulate methodologies and techniques for investigating the model.

The following are objectives of this dissertation research:

- To determine the appropriate form of model for the Internet in supporting virtual cooperative interaction.
- To characterize the processes of cooperative interactions that have evolved on the Internet.
- To analyze Miller's living systems theory for its application to the socio-technical cooperation through the Internet.
- To define the socio-technical processes underlying virtual cooperative interaction.
- To analyze the roles of awareness and its technical support in virtual cooperative interaction.
- To develop methodologies and analysis software for modeling discourse patterns and social structure in virtual cooperative interaction.
- To evaluate the methodologies and apply them to a sample virtual community.
- To develop techniques and methodologies for analyzing diffusion processes in the Internet community at large.

## *1.5 Dissertation Structure*

The rest of this dissertation is divided into two parts to address the above objectives:

I. The first part introduces and develops the cyberorganism framework for modeling virtual cooperative interaction. It investigates the socio-technical origins and evolution of the net. The nature of the net as a living system is explored and examined. It then analyzes system imperatives, structures, and processes involved in the global cyberorganism.

- Chapter 2 introduces historical background about the origins of the net and virtual cooperative interaction. It traces the following developments: (i) human-computer symbiosis; (ii) ARPANET, its development and evolution into the Internet; (iii) cultures and social norms of scientific enterprise as the social and psychological foundations for the development of the virtual cooperative interaction.

- Chapter 3 develops the cyberorganism framework further by introducing 20 critical subsystems' in living systems. Relationships among subsystems are examined. The chapter concludes with a listing of critical subsystems within the emerging global cyberorganism.

- Chapter 4 focuses further on information subsystems and processes described in Chapter 3, particularly the decider, channel and net, associator, memory, and timer subsystems. They involve awareness issues which are crucial for coordination processes. System awareness is needed for maintenance of feedback loops within the living system.

- Chapter 5 examines five key elements in virtual cooperative interaction: discourse patterns; time-dimension; awareness hierarchy; motivations for cooperative behaviors; and, emergence and maintenance of virtual cooperative interaction.

II. The second part investigates the conceptual framework's utility and application. It outlines techniques and methodologies for investigation of virtual cooperative interaction. Its demonstrations of the utility of the cyberorganism framework correspond to the three system levels in the cyberorganism: *team*, *special interest community*, and *the Internet community at large* (i.e., group, organization, and community within Miller's theory).

- Chapter 6 presents awareness support mechanisms such as CHRONO and methodological dimensions for those mechanisms. The main interests are awareness maintenance for teams and special interest communities.

- Chapter 7 describes systematic methodologies for investigating social and psychological structures in special interest communities through listserver analyses. It introduces time-series, sociometric and group-dynamic approaches in analyzing vitality, social structure, and psychological dimensions in virtual communities.

- Chapter 8 demonstrates the listserver analysis software: ListA and applies the methodologies described in Chapter 7 to a special interest community as a sample case.
- Chapter 9 specifies constructs and methodological approaches for tracking memetic diffusion and dissemination processes. It investigates the reproduction, migration and dissemination of vehicles/memes in the Internet community at large.

Finally, Chapter 10 summarizes major contributions of this dissertation, and concludes the with future research direction.

# CHAPTER 2

# The Living Internet and its Origins

In order to study and support collaborative activities on the net, we need a conceptual framework that identifies the major distinctions among "virtual groups" of widely differing sizes and structures, and among the various roles that originators, recipients, and intermediaries can play.

A living systems perspective allows us to model the net as an emergent, cybernetic organism, or *cyberorganism*. It is a global living system developing, evolving and inhabiting in information sphere or infosphere, analogous to a biological organism living inside biosphere. Using this conceptual framework, we can analyze and classify the types of information systems which support communication, social, and knowledge processes involving virtual cooperative interaction.

## *2.1 Living Systems Perspective*

The basic **living systems theory** was built upon a search for the common properties of all living systems. The theory is the life's work of James Grier Miller (Miller, 1978; Miller & Miller, 1990; Miller & Miller, 1992). It should be noted at this point that the basic theory is in flux, and will continue to be so. The theory demonstrates that living systems exist at eight levels of increasing complexity from cells through organisms, to communities, societies and supranationals (Bailey. 1994).

Living systems are concrete, open systems possessing the characteristics of life. That means they are composed primarily of sub-compounds and are generated by genetic and/or memetic templates. All living systems tend to maintain steady states (or homeostasis) of many variables, keeping an orderly balance among subsystems which process matter-energy and information. They actively regulate themselves to maintain steady states of negentropy, as well as to grow, develop, and propagate. This motivates a *teleological stance* to living systems in which they seem to strive purposefully to preserve their own system or structure against entropy, the universal tendency toward disintegration.

Those processes in subsystems which maintain steady states are *adjustment processes*. They operate based on the notion of feedback. When signals are fed back over the feedback channel in such a manner that they increase the deviation of the output from steady state, *positive feedback* exists. When the signals are reversed, so that they decrease the deviation of the output from steady state, it is *negative feedback*. Positive feedback alters variables and destroys their steady states in systems. Thus, positive feedback can initiate system changes and growth. Negative feedback maintains steady states in systems. It cancels an initial deviation or error in performance (Miller, 1978).

**Cybernetics**—the study of methods of feedback control—is an important part of the systems theory. It has led to the recognition of certain formal identities among various sorts of nonliving and living systems. In a complex system, control is achieved by many finely adjusted, interlocking processes involving transmissions of matter-energy and information (Miller, 1978). The word, *cybernetics* is derived from Greek—*kybernetes*, meaning "steersman"—which Wiener (1948) defined as the science of "*control and communication in the animal and the machine*". Originally cybernetics were concerned with patterns of communication, especially in closed loops and networks. Later investigation led to the concepts of feedback and self-regulation and then, to self-organization.

This attention to patterns of organization lead Wiener to recognize that the new notions of message, control, and feedback referred to patterns of organization—that is, to nonmaterial entities—are crucial to a full scientific description of life. Later, he expanded the concepts of pattern, from the patterns of communication and control that are common to animals and machines to the general idea of interaction pattern as a key characteristic of life: “we are but whirlpools in a river of ever-flowing water; we are not stuff that abides, but patterns that perpetuate themselves” (Wiener, 1950; Capra, 1996).

Living systems, as open systems, exhibit interaction patterns with their surrounding environment and other living systems. They must regularly acquire resources to replace those that are consumed by the transformation processes of the system or lost through dissipation and extrusion. Some resources are derived from the nonliving part of the environment, but much of the input comes from other living systems. The products or wastes extruded by one system may be valuable inputs to others (Miller, 1978).

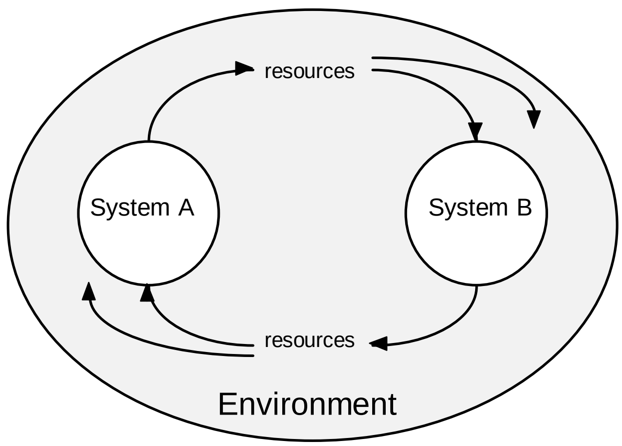


**Figure 2 Systems Dyad**

Much of virtual cooperative interaction is involved with the exchange of resources between cybernetic living systems. Such exchanges will be a major theme of this dissertation. To portray them we will employ a simple exchange model called *systems dyad* (Tracy, 1989), shown in Figure 2. This model emphasizes the fact that the typical exchange is two-way, and that other originator systems and receiver systems may exist. In some cases, the interdependence between the systems becomes so great the dyad itself becomes a supra-living system.

Establishing feedback loops between cybernetic living systems is essential in coordinating resource exchanges among them. In order for them to engage in mutual cooperative interaction, they initially need to be aware of the existence of one another.

Afterwards, they need to be aware of what each is doing in order to continue the interaction. Here *awareness mechanisms* in supra-living systems transmit coordination signals among its subsystems or components. They constitute the feedback channels and regulate the adjustment processes within the supra-system.

Living systems include all forms of animal and plant life as well as organized groupings (social systems) of organisms. Miller identified eight levels of living systems: cell, organ, organism, group, organization, society, community and supranational. In this dissertation, we will be concerned primarily with the subset of human systems and with the middle levels—individuals, groups, organizations, and communities.

### 2.1.1 Evolution of Life

Evolution, broadly speaking, is the self-organizing process, not only of life, but of the universe itself. The ordering of matter into elementary particles and then into planets and stars was a necessary prelude to life on earth. Life arose out of the conditions set by those earlier events (Hoagland & Dodson, 1995). Over the last 3.8 billion years a continuous biosocial evolution has occurred in the overall direction of increasing complexity. The living systems theory asserts that all of the great variety of living entities that evolution has produced are a complex structured open-system. They maintain, within their boundaries, their thermodynamically improbable energy states by continuous interactions with their environments. Inputs and outputs for both matter-energy and information are essential for living systems. The total inputs are lower in entropy and higher in information than the total outputs (Miller & Miller, 1990). Growth was directed by the structure of atoms and molecules that permitted some combinations and inhibited others. In this scenario we see at the elemental level two of the prerequisites of life: available matter-energy and information that provides a pattern or template for growth (Miller, 1987).

#### 2.1.1.1 Origin

Our story begins on the streaming, turbulent surface of early earth, similar to locations like the hot springs that exist today, where we find bacteria called Archaea. These organisms are known to be very ancient, and they thrive in temperatures near the boiling point of water. Nucleotides and amino acids were probably plentiful before life appeared. Not only can these essential building blocks for DNA (deoxyribonucleic acid), RNA (ribonucleic acid), and protein be made surprisingly easily and thus may have been assembled spontaneously right here on earth, but they have been found in space dust and meteorites that are likely to have showered down on earth over the eons (Hoagland & Dodson, 1995).

Once, the first reproducing nucleotide chains—probably RNA—had been formed, some of them developed a remarkable ability: they could copy themselves. They were not alive in any sense, but simply floated about the *pre-biotic soup,* mindlessly self-replicating. A self-replicating molecule needs two special properties: it must be a *template*—(1) a sequence of units (nucleotides) along which a complementary sequence of similar units can be ordered. (2) It must be an *enzyme*, able to pull free nucleotides from the surroundings and bond them together along the template. (Hoagland & Dodson, 1995).

Eventually certain molecules became so large and complex that they took on the characteristics of life. That is, they became capable of acquiring resources from their environment and building replicas of themselves. The information aspect of the molecule was then able to reproduce itself with another collection of atoms, adding structure where it did not exist before (Tracy, 1989). The molecule becomes a template for building new molecules in its own image. Thus, it becomes a *replicator* (Dawkins, 1989b; Hull, 1988). The ability to obtain resources from the environment meant that the molecule could add to its store of matter-energy in order to grow or replace losses. The molecule, or at least its information aspect, become potentially immortal, limited only by continued availability of resources and existence of favourable  conditions in the environment.

These living, reproducing molecules eventually developed ways of surrounding themselves with a captive environment and evolved into the first cells. Single-celled organisms developed into multi-celled organisms. Specialization among certain groups of cells evolved into specialized organs, permitting more elaborate organisms to develop. Cells, organs, and organisms formed three levels of living systems, each having the basic characteristics of life (Miller, 1978; Tracy, 1989). In essence, evolution proceeds by gradual tinkering. Complex living systems had, at one time, cruder and simpler predecessors. Small improvements then accumulated in such a way as to produce a big change over time. However, evolution proceeds without a foreseen purpose or direction. Random changes, cumulative selection (i.e., innovation that builds on top of prior innovations), and many successive generations are what allow evolution to work (Hoagland & Dodson, 1995).

#### 2.1.1.2 Replicator

Dawkins (1989b) depicts the development of living systems from the unique viewpoint of genes. Some basic definitions are useful here. First, a *gene is a unit of heredity*. Technically the term **gene** refers to 'any hereditary information for which there is a favorable or unfavorable selection bias equal to several or many times its rate of endogenous change'. More generally, gene means 'that which segregates and recombines with appreciable frequency' (Dawkins, 1989a). Second, **genotype** is the genetic constitution of an organism at a particular locus or set of loci. Sometimes used more loosely as the whole genetic counterpart to phenotype. Third, **phenotype** denotes attributes of an organism, the joint product of its genes and their environment during ontogeny. A gene may be said to have phenotypic expression in, say, eye colour. The concepts of phenotype is *extended* to include functionally important consequences of gene differences, outside the bodies in which the genes sit (Dawkins, 1989a).

Genes have the ability to replicate themselves and to serve as templates for the growth of surrounding systems. This ability allows genes to proliferate structures that enhance the potential for genetic survival. Thus cells, organs, and organisms may be seen as "survival machines" built by genes to cope with such environment changes as the thinning of the primordial soup and the development of predators (Tracy, 1989). They all are survival machines for the same kind of replicator—DNA molecule—but there are many different ways of making a living in the world, and the replicators have built a vast range of machines to exploit them. A monkey is a machine that preserves genes up trees, a fish is a machine that preserves genes in the water (Dawkins, 1989b).

In some ways, human culture and social behaviour can be traced to genetic survival. Even social systems such as the family and the hunting pack can be linked to the struggle of genes for survival and propagation. At the social level, however, a new element—the *meme* (Dawkins, 1989b)—is encountered. The word, **meme**, conveys the idea of a unit of cultural transmission, or a unit of *imitation* (comes from the Greek root of 'Mimeme'). Memes are self-replicating ideas. For example, concepts such as nationalism and democracy are memes, because they tend to generate social systems that pass the concept from generation to generation. But a meme can be something as simple as a catchy tune or a fashion design, so long as it induces living systems to propagate it (Dawkins 1989b; Dennett, 1995; Hull, 1988).

Memes share many similar characteristics with genes. The essence of memes, as with genes, is information (i.e., pattern or structure) that is capable of replicating itself. Memes are potentially immortal, although subject to mutation. As genes generate various patterns of life at the cellular, organic and organismic levels, memes generate a variety of cultural patterns among individuals, groups, organizations, communities and societies. Memes require survival machines or *vehicles* (Dawkins, 1989b; Hull, 1988) for their maintenance, actualization and propagation. Genes and memes form the *template* or *charter*—a set of instructions outlining the basic structure and processes of the living system. This set must exist from the moment of origin of the system, although it may subsequently be modified. Genes provide the basic template for cells, organs, and organisms. Memes and genes together provide the charters of social systems.

To the extent that common structures and functions can be identified at all levels among these living vehicles of genetic and memetic information, these common features may be regarded as the basic requisites of a living system. Hence, the living systems theory is the study of the fundamental structures and processes found in the survival machines that are generated by genes and memes (Miller 1978; Tracy 1989).

### 2.1.2 Imperatives of Life

Three broad purposes or imperatives of life are exhibited by genes, memes and their surviving vehicles—i.e., the host living systems.

The first is immediate survival of the system through *maintenance of steady states*. When opposing variables in a living system are in balance, the system is in equilibrium with regard to them. In the near term, a living system consists of a larger number of relationships or variables that must be held at, or near, steady states (Miller, 1978). For example, the human body must maintain a fairly constant heartbeat, metabolism, temperature, flow of oxygen, level of iron in the blood, and so on (Tracy, 1989). Likewise a computer network must maintain a reasonably steady bandwidth flow, processing resource, supply of electric energy, number of users, and so forth. In order to maintain such steady states under entropy conditions, the system must also regularly take in fresh resources such as eating food or adding communication bandwidths. For memes the system must also maintain memory and/or social order.

There is a range of stability for each of numerous variables in all living systems. Ordinarily, there is a standard range of rates at which input enters a system. If the input rate falls below this range, it constitutes a *lack stress*. If the input rate goes above this range, it is an *excess stress.* Systems undergo stress in various ways. One class of stress is the *matter-energy stresses*, including: (a) matter-energy input lack or underload—starvation or inadequate fuel input; (b) matter-energy input excess or overload and (c) restraint of the system, binding it physically. Another class of stress is the *information stresses*, including: (a) information input lack or underload, resulting from a dearth of information in the environment or from improper function of the external sense organs or input transducers; (b) injection of noise into the system which has an effect of information cutoff, much like the previous stress; and (c) information input excess or overload. Information stresses may involve changes in the rate of information input or in its meaning (Miller, 1978).

Adaptation to change, when the basic aim is to maintain the health and integrity of the system, falls within the maintenance imperative. For instance, the maintenance of the steady state of message flow is essential for the health and vitality of a listserver-based virtual community. A virtual community needs to monitor and regulate the information flow among its members. It has to adapt to changing patterns of topical interests and social behaviours of its members. Message overflow or underflow on a list can induce information stresses and strains to the well-being of the system.

The second imperative is *actualization of the system's potential*. Actualization generally requires both growth (i.e., incorporation of additional elements into system) and development (i.e. elaboration of the system to cope with greater complexity in the environment). The objective of this sort of change is purposeful expansion of the system's capabilities, not simply adaptation to maintain the existing system (Miller, 1978). Growth and development are long term survival strategies built into the templates of living systems. Larger and more elaborate systems tend to be better able to control their environment; they feed on simpler systems. For example, the WWW—a more elaborate system—feeds on simpler systems like FTP and Gopher. Furthermore, memes may demand actualization. Ideas such as artificial intelligence, information superhighway, capitalism and communism cannot simply remain on the drawing board; they must be tried. Some memes seem inevitably to spawn decades, even centuries of investigation, analysis, elaboration, and development into practice or hardware (Tracy, 1989).

The third imperative is *propagation of the system through reproduction and/or dissemination*. Each gene strives to perpetuate its peculiar pattern. The current survival vehicle cannot be maintained forever, but in theory, the gene or meme can. Reproduction is the ultimate mechanism for genetic survival. For memes, dissemination serves a similar function. The more widespread an idea is, the more likely it will survive (Miller, 1978; Tracy, 1989).

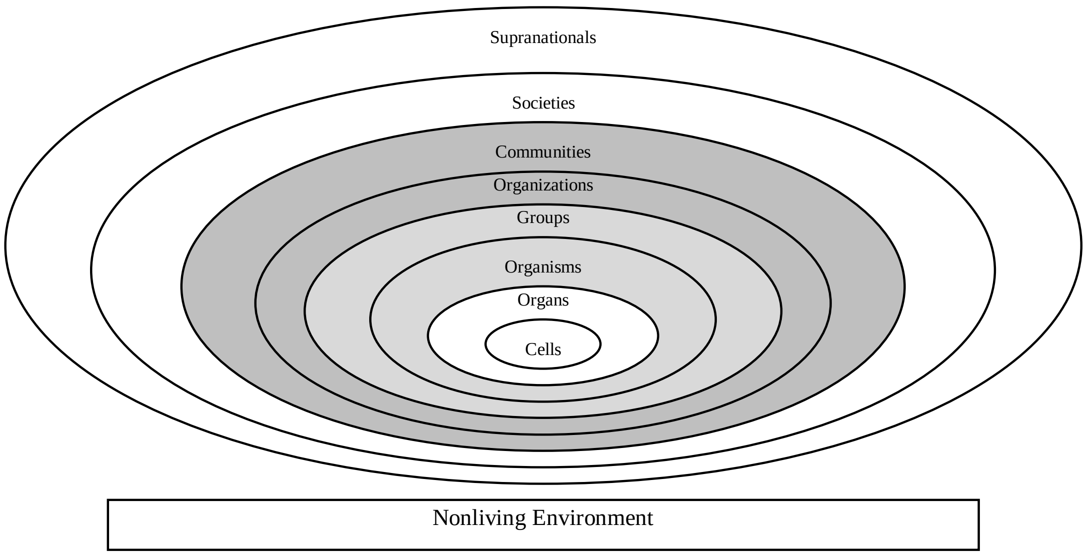


**Figure 3 Levels of Living Systems and Non-living Environment**

### 2.1.3 Living Systems Characteristics

Social systems exhibit many of the same basic characteristics as cells, organs, and organisms. Miller (1987) identified five higher levels of living systems; group, organizations, communities, societies, and supranationals. Systems at each level serve as suprasystems for systems at the next lower level. Thus, the United Nations (a supranational) is a supra-system, albeit a weak one, for nations (societies) of the Earth. The IBM Corporation (an organization) is supra-system for its employees as well as component of its suprasystems, the United States and other nations in which it operates. The eight levels of living systems and their relationships to each other and to the non-living environment are shown in Figure 3 (Miller & Miller 1990).

Suprasystems provide a certain amount of guidance and training for their component systems. For example, the family (a group) provides training for its member organisms in language, customs, morals, survival techniques, and many matters. Thus, there is a passing down of memetic heritage just as there is a downward flow of the genetic heritage (Tracy, 1989).

By means of an exhaustive review of literature in the life and social sciences—biology, botany, microbiology, physiology, zoology, psychology, sociology, anthropology, and political science—Miller was able to extract a set of general characteristics of all living systems. Living systems theory applies to persons, groups, organizations and communities because they all share certain characteristics of life (Miller; 1978; Tracy, 1989; Miller & Miller, 1990). These characteristics are listed below:

1. They are open systems, receiving inputs, processing them, and releasing outputs.
2. They are concrete systems largely composed of organic compounds, although they may include non-living components. These non-living components can be considered as *extended phenotypes* of the genes and memes (Dawkins, 1989a). Human artifacts are good examples (e.g., dental fillings, clothes, weapons, tools, buildings).
3. They posses a template from the moment of origin. The template may be encoded in genetic material (e.g., DNA, RNA), written document (e.g., constitution), or oral culture and custom (e.g., fairy tales, social norms). The template provides primary instructions for the development and functioning of the system's structures and processes. Additional instructions may come from the templates of subsystems and suprasystems.
4. They can exist only in an environment that stays within narrow ranges of temperature, air pressure, and other variable. They monitor and act upon environment to try to maintain these necessary conditions.

5. They maintain steady states of negentropy, but at the same time they tend to grow in size and complexity and to reproduce themselves or disseminate parts of their structure.
6. They are consisted of critical subsystems which carry out essential processes for survival. Smooth coordination and cooperation among those subsystems require feedback loops or/and awareness supports.
7. They each posses a decider subsystem that controls the interactions of all other subsystems and components of the system. The decider also mediate conflicting instructions from various subsystems and suprasystems.
8. They also posses or have access to other critical subsystems having functions of reproduction, matter-energy processing, or information processing. Or they have *symbiotic* or *parasitic* relationships with other living or non-living systems which carry out the processes of any such subsystem lack.
9. Their subsystems are integrated by template and decider into an actively self-regulating, developing system with purposes and goals.

We have already discussed several of these characteristics, including the fact that living systems are concrete, open systems, that they display imperatives of maintenance, actualization, and propagation. The subsystem structure of living systems will be discussed in detail in Chapter 3. Perhaps the only characteristic that requires elaboration at this point is the need for a supportive environment. Life was created originally out of an environment that provided the necessary matter and energy to sustain certain chemical reactions. Cell developed some ability to store needed resources and protect themselves from adverse conditions, and each successive level of living systems expanded these capabilities. The scientific literature on plans for deep-space exploration provides an interesting glimpse of how living systems continue to extend the range of environments within which life can exist (Tracy, 1989). Nevertheless, there are still limits to the ability of living systems to sustain themselves. Boundary and supporter subsystems break down when there is too great a differential of temperature or pressure between a system and its environment (Miller, 1978; Tracy, 1989).

Considering the characteristics of living systems as a whole, everything seems aimed at preservation and extension of the system. Living systems act to stem, and in a limited way reverse, the universal tide of entropy. They do so by maintaining, actualizing, and propagating their own particular brand of order. These are their primary imperatives, as mandated by their templates (Tracy, 1989).

To understand the origins of the Internet is to understand how it got where it is today and more important how it is poised to continue this evolution into a cybernetic living system. As stated earlier, replicators like genes and memes are the basic templates for living systems. Original memes, or ideas for building the Internet, required suitable environments and conditions. Once a *proto*-system (like ARPANET) came into existence, in addition to its need for the immediate survival through system maintenance, the second imperative of life dictated the drives for actualization of its potential through growth and development. The third imperative eventually encouraged the propagation of the network system through reproduction and/or dissemination. In the following subsections, an evolutionary story of the Internet will be told. From analyses of its origins, we will see how it grows and develops in complexity, eventually becoming a cyberorganism.

## *2.2 Genesis*

After the surprise 1957 launch of the Soviet satellite Sputnik, the US Department of Defense (DOD) established the Advanced Research Project Agency, or ARPA, to strength national security through far-researching research. Like the primordial soup for the organic replicators to come into being, ARPA together with its associated research universities and institutions became an ideal environment for memes that eventually would lead to Internet to develop and replicate (Dertouzos, 1997).

The DOD viewed the computer, which was still an emerging technology, as potentially important to military command and control. But the director of the agency's Information

Processing Technique Office, the late J. C. R. Licklider, had a broader view. A psychologist, Licklider saw a new era in which computers and people would act in concert creating a *human-computer symbiosis* (Licklider, 1960).

### 2.2.1 Human-Computer Symbiosis

Mutualistic interactions between living systems are an integral part of life. Some mutualistic relationships are so pervasive that they affect almost all life forms. Most animals rely on the microorganisms in their gut to properly digest and metabolize food. Termites require cellulose-digesting microorganisms in their gut to obtain all possible nourishment that their diet of wood can provide. In return those microorganisms receive digested wood as their foods (Wilson, 1975).

Human-computer symbiosis was a cognitive meme that "human brains and computing systems will eventually be coupled together very tightly, and that the resulting partnership will *think* as no human brain has ever thought and process data in a way not approached by the information processing systems we know today" (Licklider, 1960). Consequently this meme naturally allied itself to a companion meme—a conceptual framework called *augmented knowledge workshop* developed at Stanford Research Institute in 1962 (Engelbart, 1988). The framework's objective was to augment knowledge-work capability of human minds with new computer technology. Metaphorically, this co-meme depicted computer augmented collaborative group as a *new organism* which can achieve new levels of sensory capability, speed, power, and coordination as to become a new species. It eventually led to the development of many classical concepts in human-computer interaction, like the mouse input device, integrated text and graphics, multiple windows, and editing across windows, (Goldberg, 1988).

### 2.2.2 Anarchic Survival Meme: the RAND Proposal of 1964

Some 30 years ago, the RAND Corporation, a foremost Cold War think-tank in the United States, faced a strategic problem: how could the US authorities successfully communicate after a nuclear war? Post-nuclear America would need a command-and-control network, linked from city to city, state to state, base to base. But no matter how thoroughly that network was armored or protected, its switches and wiring would always be vulnerable to the impact of atomic bombs. A nuclear attack would reduce any conceivable network to tatters (Sterling, 1993).

And how would the network itself be commanded and controlled? Any central authority, any network central command, would be an obvious and immediate target for an enemy missile. The centre of the network would be the very first place to go. The RAND proposal was made public in 1964. In the first place, the network would have no central authority. Furthermore, it would be designed from the beginning to operate while significantly dangled. The principles were simple. The network itself would be assumed to be unreliable at all times. It would be designed from the beginning to transcend its own unreliability. All the nodes in the network would be equal in status to all other nodes, each node with its own authority to originate, pass, and receive messages. The messages themselves would be divided into packets, each packet separately addressed. Each packet would begin at some specified source node, and end at some other specified destination node. Each packet would wind its way through the network on an individual basis (Baran et al, 1964). This RAND survival meme for a decentralized network would become a catalyst for building the first proto-system: ARPANET.

### 2.2.3 ARPANET: First Four Network Nodes in December 1969

During the 60s, the Pentagon’s ARPA was sponsoring computer research at leading universities and research labs in the United States. These projects and their computers provided an ideal environment for an experimental network project (Roberts, 1988). Gradually this intriguing meme of an anarchic, survival-resilient, decentralized, packet-switching network began to circulate among researchers at RAND, MIT and UCLA. The National Physical Laboratory in Great Britain set up the first test network on these principles in 1968. Shortly afterward, the ARPA decided to fund a larger, more ambitious project in the USA. The nodes of the network were to be high-performance minicomputers at the time. These were rare and valuable machines which needed to be networked to key researchers, for the sake of national research-and-development projects.

In fall 1969, the first such node was installed in UCLA. By December 1969, there were four nodes on the infant network, which was named ARPANET, after its Pentagon sponsor. The four computers could transfer data on dedicated high-speed transmission lines. They could even be programmed remotely from the other nodes. Thanks to ARPANET, scientists and researchers could share each other's computer facilities by long-distance. This was a very valuable service, for computer-time was precious in the early 1970s. As soon as the first four nodes were brought up and tested in December 1969 the network grew very rapidly. One year later, in December 1970, the network had grown to 10 nodes and 19 host computers. By April 1971, there were 15 node with 23 host computers; and by 1972, 37 nodes (Roberts, 1988).

By the second year of operation, however, the emergent behaviours of network computing became clear. Users on the ARPANET had warped the computer-sharing network into a dedicated, high-speed, federally subsidized electronic post-office. The main traffic on ARPANET was not long-distance computing. Instead, it was news and personal messages. Researchers were using ARPANET to collaborate on projects, to trade notes on work, and eventually, just to gossip (Sterling, 1993). People had their own personal user accounts on the ARPANET computers, and their own personal addresses for electronic mail (email). Not only were they using ARPANET for person-to-person communication, but they were very enthusiastic about this particular service—far more enthusiastic than they were about long-distance computation. It wasn't long before the invention of the e-mail listserver, an ARPANET broadcasting technique in which an identical message could be sent automatically to large numbers of network subscribers.

Three key activities "led" from the ARPANET to today's Internet. The first was the formation of a chain of grass-roots groups that would steer the future Internet standard in the early 1970s. This approach marked a major break in the way standards were formed. Instead of the top-down processes that took years to arrive at a consensus, the new groups operated in an informal manner, seeking advice, trying a quick idea here, giving out some code there, to see if it "took" and until it "felt right". These seemingly ad hoc processes moved the networking effort steadily forward. Interestingly the web would follow the same path. The second was the wide adaptation of TCP/IP protocol—a method of addressing many different networks using a long number. The third key event is the emergence of local area networks (LANs), which hooked computers and workstations together within building. The LANs became possible largely because of the invention of the Ethernet. Through the rapid growth of personal computers and workstations in 1980s, LANs came into widespread use and placed a huge demand for connectivity on the burgeoning Internet (Dertouzos, 1997).

### 2.2.4 RFCs: Gametes and Zygotes for Memes in 1969

The first key to the rapid growth of the ARPANET into INTERNET has been free and open access to the basic documents, especially the specifications of the protocols. (Leiner, Cerf, Clark, Kahn, Kleinrock, Lynch, Postel, Roberts &Wolff, 1997). The beginning of the ARPANET in the university research community promoted the academic tradition of open publication of ideas and results. However the normal cycle of traditional academic publication was too formal and too slow for the dynamic exchange of ideas essential to creating networks. In 1969, a key step was taken by Crocker (1969) in establishing the request for comments (or RFC) series of notes. These memos were intended to be an informal fast means of distribution for sharing memes among network researchers. At first the RFCs were printed on paper and distributed via postal mail. As the File Transfer Protocol came into use, the RFCs were prepared as online files and accessed via FTP. SRI (Stanford Research Institute), in its role as Network Information Centre, maintained the online directories.

RFCs are like *gametes* for memes. A **gamete** is a sexual cell capable of fusing with another in reproduction. The circulation effect of the RFCs was to create a positive feedback loop, so memes or proposals presented in one RFC would trigger other RFCs. Such a series of memetic mixing processes is analogous to genetic recombination in sexual reproduction. It is complex and remarkably uniform in its objective for providing evolutionary plasticity. Biologists very commonly regard sexual reproduction as a biotic adaptation. The general machinery of sexual production is directed at the goal of *producing, with the genes of two (or more) parental vehicles, offspring of diverse genotypes* (Williams, 1966). Likewise when consensus (a stable recombined set of memes) would come together, a specification document would be prepared: such specifications would then be used as the basis for implementations by the various research teams. Specifications resulting from RFCs are now viewed as the "documents of record" (Leiner et al, 1997) in the Internet engineering and standard community. RFCs will continue to be critical to net evolution while furthering the net's initial role of sharing information about its own design and operations. Like a **zygote**—the cell formed by the union of gametes; the first cell in development of a new organism—for genes, a resulting specification becomes the *de facto* zygote for memetic transmission and development.

### 2.2.5 TCP/IP Network Protocol: New Development in 1977

Throughout the 1970s, ARPA's network grew. Its decentralized structure made expansion easy. Unlike standard corporate computer networks, the ARPA network could accommodate many different kinds of machines. As long as individual machines could speak the packet-switching *lingua franca* of the new, anarchic network, their architecture, and their operation system (OS), and even their ownership, were irrelevant.

The ARPA's original standard for communication was known as NCP, "Network Control Protocol," but as time passed and the technique advanced, NCP was superseded by a higher-level, more sophisticated standard known as TCP/IP. TCP, or "Transmission Control Protocol," converts messages into streams of packets at the source, then reassembles them back into messages at the destination. IP (Internet Protocol ), handles the addressing, seeing to it that packets are routed across multiple nodes and even across multiple networks with multiple standards —not only ARPA's pioneering NCP standard, but others like Ethernet, FDDI, and X.25.

As early as 1977, TCP/IP was being used by other networks to link to ARPANET. ARPANET itself remained fairly tightly controlled, at least until 1983, when its military segment broke off and became MILNET (Sterling, 1993). But TCP/IP linked them all. And ARPANET itself, though it was growing, became a smaller and smaller neighborhood amid the vastly growing network of other linked machines.

### 2.2.6 Birth of the Internet: Dissemination/Reproduction in 1984

As the ‘70s and ‘80s advanced, many very different social groups found themselves in possession of powerful computers. It was fairly easy to link these computers to the growing network-of-networks. As the use of TCP/IP became more common, other entire networks fell into the digital embrace of the Internet, and messily adhered. Since TCP/IP was public-domain, and the basic technology was decentralized and rather anarchic by its very nature, it was difficult to prohibit people from linking up somewhere-or-other. In point of fact, no one wanted to stop others from joining this branching complex of networks, which came to be known as the Internet (Sterling, 1993). Connecting to the Internet cost the taxpayer little or nothing, since each node was independent, and had to handle its own financing and its own technical requirements. Like the phone network, the computer network became steadily more valuable as it embraced larger and larger territories of people and resources.

In 1986, the US National Science Foundation (NSF) initiated the development of the NSFNET which provided a major backbone communication service for the Internet. With its 45 megabit per second facilities, the NSFNET carried on the order of 12 billion packets per month between the networks it links. The National Aeronautics and Space Administration (NASA) and the US Department of Energy contributed additional backbone facilities in the form of the NSINET and ESNET respectively. In Europe, major international backbones such as NORDUNET and others provide connectivity to over one hundred thousand computers on a large number of networks (Williams, 1996). Commercial network providers in the US and Europe are beginning to offer Internet backbone and access support on a competitive basis to any interested parties.

The use of TCP/IP standards for computer networking is now global. “Regional” support for the Internet is provided by various consortium networks and “local” support is provided through each of the research and educational institutions. Much of this support has come from the federal and provincial governments, but a considerable contribution has been made by industry. In Europe and elsewhere, support arises from cooperative international efforts and through national research organizations. During the course of its evolution, particularly after 1989, the Internet system began to integrate support for other protocol suites into its basic networking fabric. The present emphasis in the system is on multi-protocol inter-working, and in particular, with the integration of the Open Systems Interconnection (OSI) protocols into the architecture.

Both public domain and commercial implementations of the roughly 100 protocols of TCP/IP protocol suite became available in the 1980’s. During the early 1990’s, OSI protocol implementations also became available and, by the end of 1991, the Internet has grown to include some 5,000 networks in over three dozen countries, serving over 700,000 host computers used by over 4,000,000 people (Williams, 1996).

### 2.2.7 Conception of the World Wide Web in 1989

The World Wide Web was conceived by Berners-Lee in March 1989 (CERN, 1994) as a "hypertext project" to organize documents at CERN in an information retrieval system (Berners-Lee and Cailliau, 1990). The design involved: a simple hypertext markup language; distributed servers running on machines anywhere on the network; and access through any terminal, even line mode browsers. The web today still conforms to this basic model. Major usage began to grow with the February 1993 release of Andreessen’s (1993) Mosaic for X-Windows. Whereas the original web proposal specifically states it will not aim to "do research into fancy multimedia facilities such as sound and video" (Berners-Lee and Cailliau, 1990), the HTTP protocol for document transmission was designed to be content neutral and as well-suited to multimedia material as to text. The availability of the rich X-Windows graphic user interface on workstations supporting color graphics and sound led naturally to multimedia support, although the initial objective of meaningful access through any terminal was retained. Much web material can still be browsed effectively through a line mode browser.

In March 1993 the web was still being presented (Berners-Lee, 1993) as primarily a hypermedia retrieval system, but in November that year a development took place that so changed the nature of the web as to constitute a major new invention in its own right. Andreessen (1993) issued NCSA Mosaic version 2 using Standard Generalized Markup Language (SGML) tags (Goldfarb, 1990) to encode definitions of Motif widgets embedded within a hypermedia document, and allowed the state of those widgets within the client to be transmitted to the server. Suddenly the web protocols transcended their original conception to support graphic users interfaces providing access to interactive, distributed, client-server information systems (Rice, Farquhar, Piernot and Gruber, 1996). This change was again serendipitous since the original objective of the design had been to enable the user to specify retrieval information in a dialog box that was embedded in a document rather than in a separate window. However, the solution generalized from an embedded dialog box to any Motif widget including buttons, check boxes and pop-up menus. The capability of the user to use a web document to communicate with computer services allows *active documents* to be published on the web that, for example, provide data analysis, animation and simulation, and hence offer major new capabilities for scholarly communication (Gaines, Chen & Shaw, 1997).

### 2.2.8 Growth and Transformation of the Internet in 1990s

The exponential growth to ubiquity of access of the net and web are quantitative measures of the fundamental utility of the services provided, especially when one notes it was that growth that led to public awareness of the significance of the information highway. Internet access and usage has grown because it satisfies a need—there was very little ‘marketing’ of the services during the basic growth period (Gaines, Chen & Shaw, 1997).

In the 1990s, the Internet continues to grow at exponential rates. Some estimates are that the volume of messages transferred through the net grows 20 percent a month. In response, government and other users have tried in recent years to expand the net itself. Once, the main NSFNET backbone in the US moved data at 1.5 million bits per second. That proved too slow for the ever increasing amounts of data being sent over it, and in recent years the maximum speed was increased to 15 million and then 45 million bits per second.

Another major change has been the development of commercial providers like Merit that provide inter-networking services at speeds comparable to those of the government system. ARPANET itself formally expired in 1989, a happy victim of its own overwhelming success. Its users scarcely noticed, for ARPANET's functions not only continued but steadily improved. In fact, by mid-1994, the US government had removed itself from any day-to-day control over the workings of the net, as regional and national providers continue to expand (Gaffin & Heitkötter, 1994).

The original usage of ARPANET by the scientific and engineering communities grew through the 1970s and in 1984 the National Science Foundation in the USA funded a program to create a national academic infrastructure connecting university computers in a network, NSFNET. In 1987 the net had grown to such an extent that NSF subcontracted its operation to Merit and other commercial providers, and in 1993/1994 the network was privatized and its operation taken over by a group of commercial service providers. Email on the Internet commenced in 1972, news distribution in 1979, gopher in 1991, and web browsers with multimedia capabilities in 1993 (Gaines, Chen & Shaw, 1997).

An October 1995 survey (CommerceNet, 1995) estimates that 8% (18M) people aged 16 or above in North America had used World Wide Web in the previous 3 months. The *Lycos* (1995) search robot had indexed 10.75M documents in October 1995 which was estimated to be 91% of the total web corpus, and the overall growth rate of documents published on the web was over 1000% in 1995. The growth rate of overall Internet traffic is some 100% a year. However, web traffic was growing at some 1,000% a year when last accurately measured using NSFNET statistics for 1993/94. The growth of web traffic is widely recognized as a major impediment to its effective application, and a number of commercial services developed to operate through the web have been discontinued because the current infrastructure cannot sustain the traffic (Bayers, 1996).

The growth of the web relative to all the other services is apparent if one plots the proportion of the data accounted for by each service. The Merit statistics can be used through to their termination in April 1995 on the assumption that the relative traffic on the original backbone is representative of that on the whole Internet after November 1994. Figure 4 shows the proportion of FTP, web (HTTP), Gopher, News, Mail, Telnet, IRC and DNS data on the NSFNET backbone from December 1992 through April 1995. It can be seen that the proportion of all services except FTP and HTTP remain relatively constant throughout the period, declining slightly towards the end. However, the proportion attributable to FTP decreases while that due to the web HTTP protocol increases and becomes greater than that through: IRC in October 1993; Gopher in March 1994; mail in July 1994; news in November 1994; and FTP in March 1995. This corresponds to the basic web protocol becoming the primary carrier of net data traffic with a 25% and growing share when last measurable.

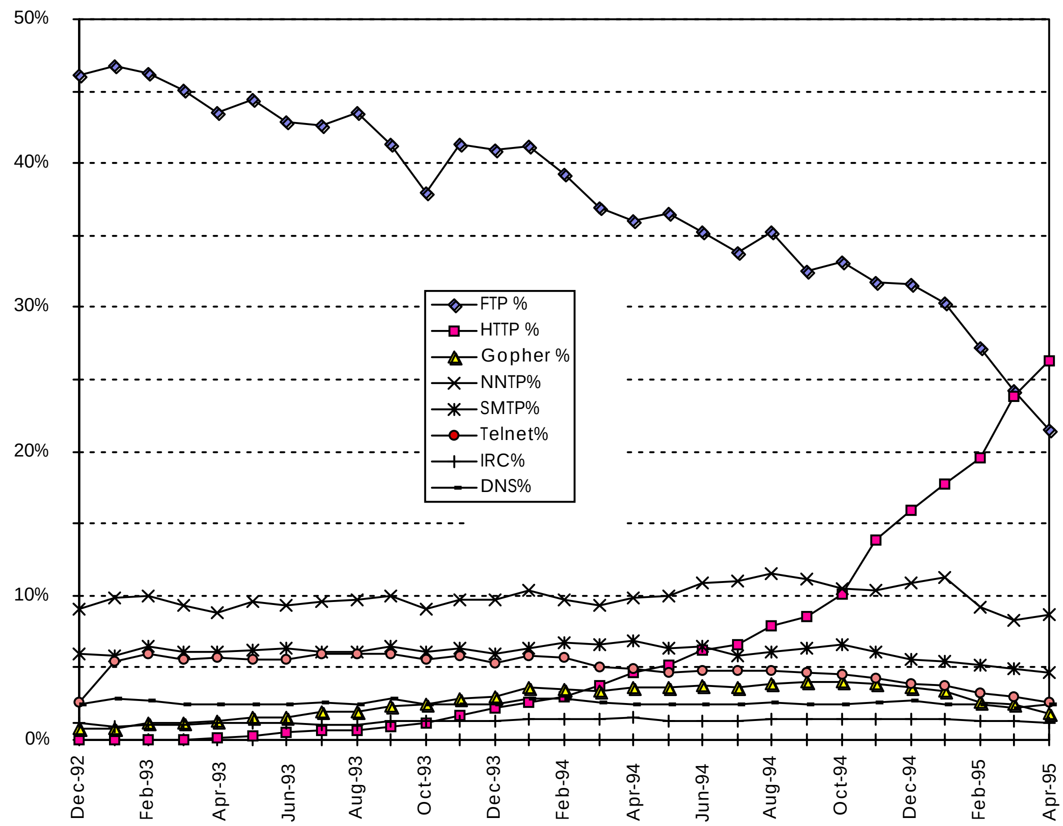


**Figure 4 Proportion of FTP, web (HTTP), Gopher, News (NNTP), Mail (SMTP), Telnet, IRC and DNS traffic on the NSFNET backbone 1992-1995**

It should be noted that one factor in the growth relative to other services is that the web traffic consists of large documents with embedded graphics. These statistics do *not* indicate that the number of web transactions exceeds the number of email transactions. It should also be noted that web browsers typically support many of the protocols shown including FTP, Gopher and News, but their usage of these protocols will show up under those protocols in the statistics. The crossover of web and FTP curves in Figure 2 shows a transition in the *servers* being primarily accessed, from FTP servers to web HTTP servers (Gaines, Chen & Shaw, 1997).

In 1971, a mere twenty-six years ago, there were only four nodes in the ARPANET network. The growth to over one million nodes, the growing commercial usage of Internet services, and the multimedia capabilities of the web in the 1993/1994 period combined to persuade government and industry that the Internet was a new commercial force comparable to the telephone and television industries, and metaphors such as *information superhighway*, *digital library*, *electronic market place*, *digital world* and *cyberspace* (Stefik, 1996) came into widespread use.

## *2.3 Emergence of CyberOrganism*

Much of the information found and retrieved from the net is generated as needed through discourse on listservers—the Internet is a mixed community of publications and intelligent human agents that both stores knowledge and generates it on demand. When the information needed cannot be found through retrieval then it may be requested through discourse, a phenomenon prophesied in the early days of timeshared computing:

> *"No company offering time-shared computer services has yet taken advantage of the communion possible between all users of the machine...If fifty percent of the world's population are connected through terminals, then questions from one location may be answered not by access to an internal data-base but by routing them to users elsewhere—who better to answer a question on abstruse Chinese history than an abstruse Chinese historian."* (Gaines, 1971)

The community of distributed intelligent agents that is the living Internet provides an 'expert system' with a scope and scale well beyond that yet conceivable with computer-based systems alone. Computer-based discovery, indexing and retrieval systems have a major role to play in that community, but are only one aspect of Internet information systems.

Krol (1993) captures the essence of these considerations in Internet RFC1462 which replies to the question "What is the Internet" with three definitions:

1 a network of networks based on the TCP/IP protocols,

2 a community of people who use and develop those networks,

3 a collection of resources that can be reached from those networks.

These are complementary perspectives on the net in terms of its technological infrastructure, its communities of users, and their access to resources, respectively. Models of computer-mediated communication must take into account all three perspectives: how agents interface to the network; how discourse occurs within communities; and how resources are discovered and accessed (Gaines, Chen & Shaw, 1997). Those integrative perspectives foretell the evolutionary direction of human-computer symbiosis with respect to cyberspace:

> ***Cyberspace****: a new universe, a parallel universe created and sustained by the world's computers and communication lines. A world in which the global traffic of knowledge, secrets, measurements, indicators, entertainment and alter-human agency takes on form: sights, sounds, presence never seen on the surface of the earth (Benedikt, 1991).*

Cyberspace is a word coined by William Gibson (1994) that gives a name to a new stage of infosphere, a new and irresistible development in the elaboration of human culture and actualization. Networked computers are the enabling media which extends human senses in this emergent virtual world (McLuhan 1964; McLuhan & McLuhan, 1988). Taking a *collective stance* (Gaines, 1994b) to Krol's (1993) definitions, we can consider the Internet as a whole as an emergent, global **cyberorganism** which develops, lives, and inhabits in cyberspace.

The net as the global cyberorganism may lead to an ultimate form of human-computer symbiosis every much like the *emergent of eukaryotic cells* (those cells having a well-defined nucleus and which higher plants and animals). Within each human cell there are numerous tiny bodies called mitochondria. The mitochondria are chemical factories, responsible for providing most of the energy we need (Dawkins, 1989b). The mitochondria that occur in eukaryotic cells are thought to have originated as separate organisms[2] that took up residence inside other cells (such as human cells). The symbiotic

---

[2] Mitochondria have their own separated DNA strains within eukaryotic cells.

origin of mitochondria in human cells had been well documented (Margulis, 1981; Dawkins, 1989b; Dennett, 1985). Eventually neither living system was able to survive without the other—a situation called *obligative symbiosis.* Similarly without network computers we have no access into cyberspace; and the network computers need us to provide survival vehicles for memes to flourish in cyberspace.

Imagine a systems dyad (Figure 2) of two different types of systems: human and computer. The systems dyad of an individual using a network computing device (e.g., network workstation, mobile digital assistant) can be visualized as an *eukaryotic* '*cybercell'* within the cyberorganism. Online virtual communities, in this respect, can be considered as 'cyberorgans' made from organizations of cybercells. Hence the Internet then becomes a global cyberorganism growing and flourishing inside cyberspace. It becomes the primary *survival vehicle* for carrying on *memes* of what philosopher Sir Karl Popper (1972) terms "*World 3*" objects, the expressed products of the human mind that continue to exist independently of their originators.

In this paradigm, the world as a whole consists of three, interconnected worlds: *World 1*, the objective world of material, natural things and their physical properties; *World 2*, the subjective world of consciousness—- with intentions, calculations, feelings, thoughts, dreams, memories, and so, in individual minds; *World 3*, the world of objective, real, and public structures which are the *not-necessarily-intentional* products of the minds of living creatures, interacting with other and with the natural *World 1* (Popper, 1972). Anthills, birds' nests, beavers' dams, and similar highly complicated structures built by animals to deal with the environment, are forerunners. But many *World 3* structures, Popper noted, are abstract: they are purely informational (e.g., forms of social organization, pattern of communication). Thus language, mathematics, law, religion, philosophy, arts, the sciences, and institutions of all kinds are all edifices of sort, like the libraries we build, physically, to store their operating instructions, their "programs" (Benedikt, 1991). Mankind's developing belief in, and effective behaviour with respect to, the objective existence of *World 3* entities and spaces meant that we could examine them, evaluate, criticize, extend, explore, and make discoveries in them, *in public*. They could evolve just as natural things do. In short, books, video tapes, CDs, libraries, temples, cathedrals, marketplaces, courts are all physical manifestations—or should we say, the physical components of—objects that exist more wholly in *World 3*. They are “objects” which are patterns of ideas, images, sounds, stories, patterns of pure information—i.e., meme. And cyberspace is the latest frontier of *World 3*.

One characteristic of living systems is that they can exist only in an environment with suitable conditions for its survival. The global cyberorganism, so called the Internet, is the emergent, living system which lives, evolves, and inhabits in this *World 3* environment called cyberspace.

### 2.3.1 Physiology of Cyberorganism: a Taxonomy of Internet Services

Physiology is the systematic study of the bodily functions of living organisms and their parts. In examining the functionality of the a living system it is useful to classify its major components of its subsystems in terms of the significant distinctions that determine their relative utilities. A general taxonomy of Internet services thus characterizes the major net services in terms their utility for computer-mediated communication, access to services or search.

It is tempting to consider the net as a new publication medium in which electronic documents emulate paper ones, and where the basic human factor issues are those of indexing and information retrieval. This makes the vast existing literature on information retrieval, its techniques and human factors, relevant to the net. However, this addresses only one aspect of computer-mediated communication, neglecting its function of supporting discourse within communities.

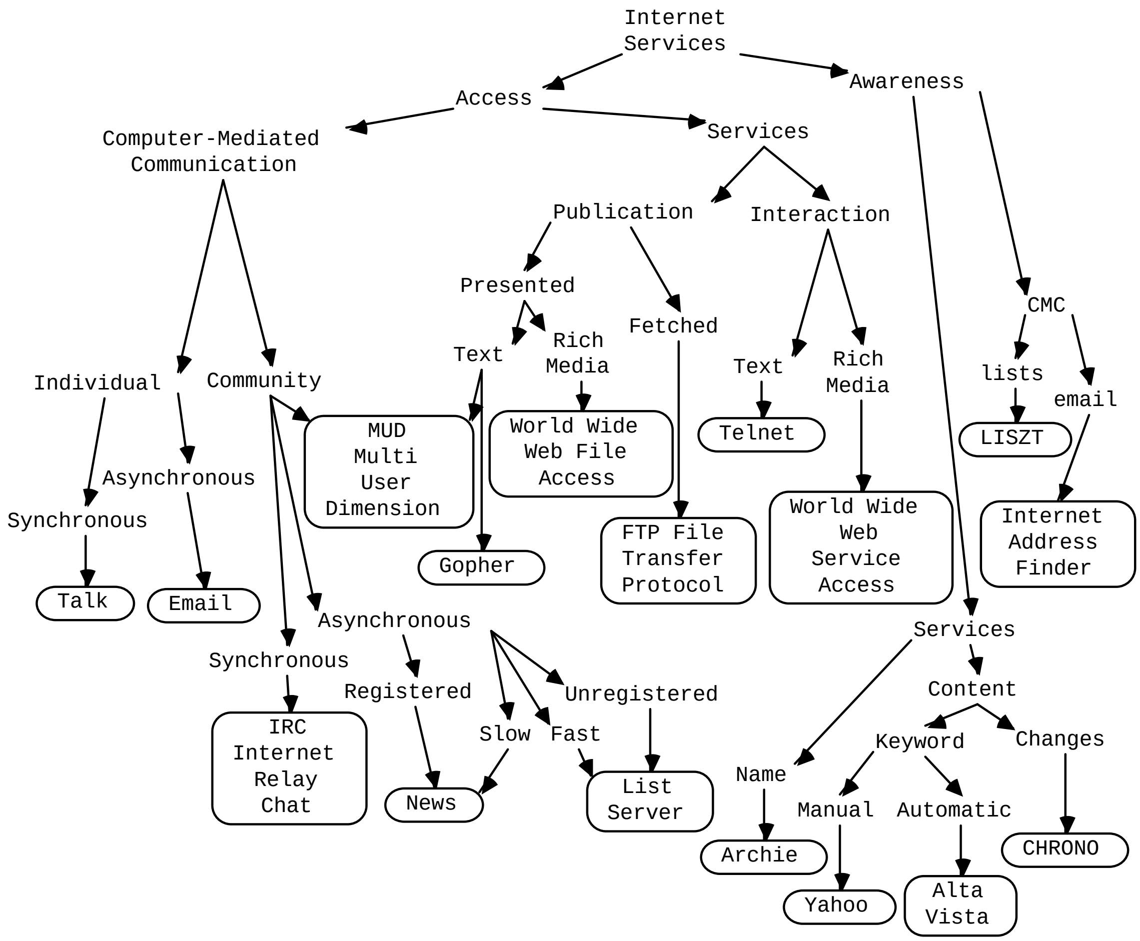


**Figure 5 General Taxonomy of Internet Services**

The taxonomy sub-classifies such communication in terms of whether it is individual-to-individual discourse or community discourse; synchronous with the participants conversing in real time or asynchronous with substantial time delays in responses. It sub-classifies asynchronous community discourse by whether the channel is slow or fast, and whether the community is centrally registered or not. It sub-classifies service access in terms of whether it is: publication or interaction; presented or just fetched; text or rich media. It sub-classifies search by whether it is: indexing communication or services; by resource name or content; by keywords or by change in contents; and whether index terms are generated manually or automatically (Gaines, Chen & Shaw, 1997).

Below are the classified components presented in Figure 5:

- *Talk*, the facility for one user to send a message directly to the terminal of another user. This provides individual, synchronous computer-mediated communication.
- *Email*, the facility for one user to send a message to the mailbox of another user. This provides individual, asynchronous computer-mediated communication.
- *Internet Relay Chat (IRC)*, the facility for a user to join a chat group and send a message directly to the terminals of the group. This provides community, synchronous computer-mediated communication.
- *News*, the facility for a user to mail a message to a registered newsgroup archive and to access messages in the archive. This provides community, asynchronous computer-mediated communication. Because the archives are maintained on a local server and updated through a chain of servers the updating is slow, possibly taking several days.
- *Listserver*, the facility for a user subscribe to a listserver and mail a message to it which it mails to all members on the list. This again provides community, asynchronous computer-mediated communication. Because the mailing to the list is fast (except for moderated groups where the mail is manually checked), listservers provide more interactive discourse than newsgroups. However, the registration of newsgroups makes them easier to discover, and, for high-volume discourse, users may prefer that it is not posted to their mailbox.
- *Multi-User Dimension (MUD)*, the facility for a user to ‘enter a dimension’, communicate directly with others there, and leave and retrieve documents. This provides community, computer-mediated communication and text resource access.
- *Gopher*, the facility for a user to retrieve a text document from a hierarchically structured archive.

- *World Wide Web file access*, the facility for a user to retrieve multi-media documents from an archive through hypertext links embedded in the document.
- *Telnet*, the facility for a user to interact with a remote machine through a console window providing a command line interface. This provides remote interactive access to services providing textual interaction.
- *World Wide Web service access*, the facility for a user to enter information into an HTML form and transmit it to a remote server. This provides remote interactive access to services providing rich media interaction.
- *File Transfer Process (FTP)*, the facility for a user to retrieve a file by site and name. This provides general file access but FTP clients generally lack the capability to present the files retrieved.
- *Internet Address Finder*, a service for a user to search for the email address of a person by their name. One of the problems of the net is the lack of an overall directory of users.
- *LISZT*, a service for a user to search for a listserver by its name. This attempts to overcome the problem that there is no central directory of listservers.
- *Archie*, a facility for a user to search the net for files by name. This provides search facilities for files with known names.
- *Yahoo*, a facility for a user to search the net for resources by name and key word through a manually entered classification. This provides search facilities for resources specified by their name or type.
- *Alta Vista*, a facility for a user to search the net for resources by content. This provides search facilities for resources with specified content.
- *CHRONO*, a facility for a user to search a site through a list of changed resources in reverse chronological order. This provides search facilities for resources by recency.

In summary, the taxonomy in Figure 5 presenting the major services on the net in terms of a small set of fundamental distinctions:

- At the top level, the major net services are characterized in terms of their utility for access to resources or awareness of resources.
- Access is sub-classified as to discourse, publications or services.
- Discourse is sub-classified by whether it is:-
  - — agent-to-agent discourse or community discourse;
  - — synchronous with the agents conversing in real time or asynchronous with substantial time delays in responses.
- Asynchronous community discourse is sub-classified by whether the channel is slow or fast, and whether the community is centrally registered or not.
- Publications are sub-classified by whether they are:-
  - — just fetched or presented when fetched;
  - — text or rich media.
- Services are sub-classified by whether they are text or rich media.
- Resource awareness is sub-classified by whether it is:
  - — by resource name or content;
  - — by keywords or by change in contents;
  - — by keywords generated manually or automatically.

The general taxonomy described above is not exhaustive; it only serves as a generative set of fundamental attributes. Since the net is constantly evolving, the net services depicted in Figure 5 should be considered as examples. New net services such as organizational conferencing mechanisms in *Lotus Notes* and *BSCW* can be considered as asynchronous systems in supporting community-oriented CMC. Similarly, recent “virtual worlds” which utilize *avatars* (i.e., 2D/3D virtual representations of participants) can be considered as multimedia extensions for MUD (synchronous community interaction). *ActiveWorlds* and *Palace* are good examples.

### *2.4 Sociality and Knowledge Creativity within CyberOrganism*

As stated earlier, the Internet not only is a network of networks based on the TCP/IP protocols, but also a community of people who use and develop those networks. In addition to the functional/structural perspectives given by the general taxonomy, the cyberorganism framework needs to address the origin and development of social norms and structures in this networked community. Chapter 1 briefly touched the topic of virtual cooperative interaction (which we will examine further in the subsequent chapters). In this section, we will explore the deep roots of virtual cooperative interaction within the cyberorganism.

One manifestation of virtual cooperative interaction on the net is that of *homepage*. The proliferation of personal homepages with cross-linkage of webpages by people who share common interests has made the exploration process on the web (i.e., net surfing) a *social experience*. Such a seemingly intrinsic rewarding experience can often be characterized as serendipitous and not necessarily task-oriented (as in traditional groupware). Through homepages, individuals create their own *virtual persona* on the web without any awareness of whom their eventual audience might actually be (i.e. without *extensional awareness* of particular recipients). However they often have a sense of who the potential audience might be (i.e. with *intensional awareness* of the type of recipient).

Sometimes individuals provide information resources to the web as a by-product during some self organization processes of their own knowledge. This type of unintentional *virtual* contribution typified the *World 3*: the world of objective and public structures which are the *not-necessarily-intentional* products of the minds (Popper, 1972). As observed earlier during our discussion of virtual cooperative interaction, this form of apparently cooperative behaviour is prevalent on the web. Why do such pro-social behaviours exist on the net? And how did it come into existence in the first place? How does a knowledge resource grow from contributions of people who share particular common interests? Also what is the growth pattern of this knowledge resource?

At the beginning of this chapter we saw the origin and evolution of the Internet. Its deep roots were the collaboration among scientific and research communities. Researchers and scientists were the originators of the ARPANET, NSFNET, and eventually the INTERNET. They established and built the core foundations for the net not only in the technical sense, but also in the cultural sense. They seeded the original memes and became the social norms of the net culture. The Internet community as a whole consists of social systems that are themselves living systems. By tracing the social value systems of the scientific enterprise, we can better understand the motivation and reinforcement processes involved in virtual cooperative interaction. They are key adjustment processes provide control and feedback loops in maintain the steady states of living systems.

### 2.4.1 Social Norms of Science: Origins of the Net Culture

Sociology of science provides both the historical background and the nature of collaborative ethos for the scientific enterprise. They are the deep roots of the net culture.

Blume (1974) has examined *social systems* of science. Sociologists use the term, *social system*, to stress *pattern* of interaction between individual characteristics of the system in question, recognizable, and not subject to unanticipated change. Maintenance of these characteristic patterns of actions is inductive of their mutually rewarding nature for participants. The permanence of social systems and of economic, religious and other subsystems indicated that the rewards received by participants in response to prescribed behaviour are unchangingly desirable and obtainable. Some sociologists have emphasized the "*exchange*" element in the operation of social systems as depicted by the *systems dyad* in Figure 2. The *information exchange* process can also be viewed in terms of "communication" and "coordination". Hence the notion of responsibility for maintaining communication becomes important. For example, a researcher needing some information has to determine if the effort required to find out if that information is on the Internet is greater than or less than the effort required to find it in other sources or to regenerate it. Similarly, a provider of information resource has to determine how much effort should be put into making it accessible to others for the dissemination process.

Thus the question becomes: what is the precise nature of the commodity of science: what do scientists contribute to the scientific community, for which they expect reward?

### 2.4.2 Values, Motivations and Reinforcements

Some sociologists view the contribution of information as crucial, although such information must be original for the higher rewards. Merton (1957) emphasized the importance of originality and the significance of establishing one's own priority in making a discovery. Because originality alone is rewarded, priority disputes are necessary for maintenance of the scientific system. They also produce neurotic anxiety in scientists, and deviant patterns of secretive behaviour (Merton, 1969; 1973).

The sociology of science is sometimes defined as a part of the sociology of knowledge, and yet the multifaceted problem of the relationship between knowledge and reality is a more general one, at the heart of the larger part of sociology. In Merton's (1973) work on the *ethos of science*—"the emotionally toned complex of rules and presuppositions that are held to be binding upon the scientist", the focus has shifted to the explicit concern of science as a social institution rather than a type of knowledge. He describes the normative notion of science as (i) a set of characteristic methods by means of which knowledge is certified; (ii) a stock of accumulated knowledge stemming from the application of these methods; (iii) a set of cultural values and mores governing the activities termed scientific; or (iv) any combination of the forgoing (Merton, 1942).

The institutional goal of science is therefore the extension of certified knowledge. The technical methods employed toward this end provide the relevant definition of knowledge: empirically confirmed and logically consistent statements of regularities (i.e., predictions). The institutional imperatives (mores) derive from the goal and the methods. And four sets of institutional imperatives—universalism, communism, disinterestedness, organized skepticism—are taken to comprise the ethos of modern science (Merton, 1942). The basic idea of interaction between the normative structure and the reward structure of science provides a suitable foundation for the understanding of science as a social institution. Such interaction results in: (i) the vigorous competition between scientists for the recognition of priority for scientific discovery (Merton, 1957; Merton, 1969); and (ii) "the Matthew Effect", that is, the enhancement of the position of already eminent scientists who are given disproportionate credit in cases of collaboration or of independent multiple discoveries (Merton, 1968). The latter phenomenon is analogous to the positive feedback loop in cybernetics (Gaines, 1996). It is a kind of snowballing effect that promotes further growth in already successful living systems.

One major institutional device for the competent appraisal of the quality of scientific work is the referee system (which derived from the norm of universalism). By tracing the evolution of the referee system from its origin, a better understanding of the process institutionalization in science can be achieved. Various factors are involved in the referee processes, such as: evaluative behaviour of editors and referees; status differences in submission of manuscripts and in rates of acceptance; patters of allocation to judge; and age structure in science (Zuckerman & Merton, 1971; Zuckerman & Merton, 1972). Those institutional factors must be taken in to account if a new information technology wishes to find wide acceptance by scientists and scholars.

### 2.4.3 Knowledge Growth and Social Network

In terms of the social dynamics involved in scientific growth, Kuhn's (1962) analysis of scientific change combines periods of continued cumulative growth (normal science) with periods of crisis or revolution. In this view, "normal" scientific activity in a research area is guided by a paradigm that defines the fundamental problems. The attention of scientists is directed toward these problems exclusively. As a result, scientific knowledge grows in a systematic fashion, building upon previous work. In times, however phenomena that the paradigm cannot explain become increasingly important. When these anomalies can no longer be ignored, the field goes through a period of crisis while the old paradigm is under attack and a new one is sought. A new paradigm is generally resisted, particularly by older scientists, until it has proven its superiority.

The ‘normal’ knowledge accumulation growth pattern in a scientific discipline is that of *logistic curve model,* commonly found in evolutionary and population biology (Boyd & Richardson, 1985; Brown & Rothery, 1993). The curve is a differential equation expressed in Equation 1: where *N* is the population size; *r* is the rate of growth; *t* is the time; and *K* is the theoretical *carrying capacity* of the environment (Gotelli, 1995). *K* encompasses many potentially limiting resources, including the availability of space, food, shelter (Hastings, 1997). In the case of scientific growth, *carrying capacity* represents the limits of generative power and explanation capability of a discipline’s current paradigm.

$$\frac{dN}{dt} = rN\left(1 - \frac{N}{K}\right)$$

**Equation 1 Logistic Growth Equation**

After solving the equation, the logistic growth equation can be expressed as Equation 2, where $N_0$ is initial population size:

$$N_t = \frac{K}{1 + \left(\frac{K - N_0}{N_0}\right) e^{-rt}}$$

**Equation 2 Population Size as a Function of Time**

For example, a typical scientific discipline or field has a cumulative growth pattern of publications as depicted in Figure 6 (Crane, 1972). It is represented as a logistic growth curve with cumulative population (i.e., number of publications) versus time increases in a characteristic S-shape fashion with the publications approaching the carrying capacity (the generative limited of the current paradigm in the discipline).

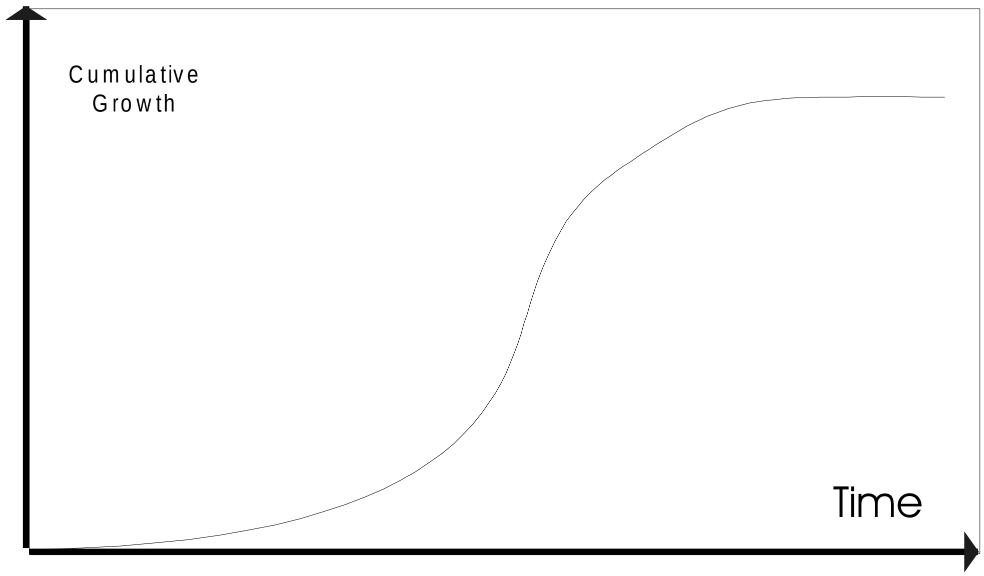


**Figure 6 Logistics Growth Curve**

It has been shown that exponential growth of a scientific research area reflects a social interaction process in which contact between scientists contribute to the cumulative growth of knowledge. The shift to exponential growth is marked by the appearance of new groups of scientists within the research area. These groups of scientists can be described from two points of view: (i) their structural aspects: who is linked to whom by sociometric ties; (ii) their normative aspects: the types of attitudes and behaviour that are expected of their members (Crane, 1972).

There are two distinct types of subgroups. One type consists of groups of collaborators. For instance, the entire set of scientists in a field may form numerous non-intersecting subgroups of various sizes. The larger ones contain a few very productive scientists and many relative unproductive ones (based on published papers). Studies have shown that these groups are linked to one another through their leaders who communicate with each other and transmit information informally across the whole field (Mullins, 1968). This enable them to monitor the rapidly changing research "front" and to keep up with new findings during a period of rapid growth. Thus the second type of subgroups in a research area is a communication network or "*invisible college*" that link groups of collaborators

(Crane, 1972). This “invisible” communication network is informal in its nature that is contrast to the formal communication channel such as the refereed system. It appears that the large groups of collaborators and the communication network that link them play very different roles in the development of the area. Under the leadership of one or two scientists, the groups of collaborators recruit and socialize new members and maintain a sense of commitment to the area among members, thus the formation of “solidarity groups” (Mullins 1968).

Concern with the speed and efficiency of scientific communication among scientists has produced a number of innovations designed to accomplish these goals. Some involve reorganizing the communication pattern within research areas: (i) changes in some aspect of the formal communication system, such as the creation of a new type of communication outlet or information service, including the replacement of the informal circulation of papers in advance of publication by a formal system that would accomplish the same purpose; (ii) improvements in arrangements for oral communications; (iii) replacement of the formal circulation if papers in “pack-ages” in the form of journals by a system of selective dissemination tailored to the needs of the individual scientist; and (iv) devices to support the scientist’s personal search of the literature such as by computerized information retrieval system. The first three types of innovations are primarily designed to increase the “visibility” of materials in the scientist’s own research areas. The fourth would aid the scientist locating materials in other areas (Crane, 1972). The development of RFC, email discourse, FTP, and WWW on the net has indeed met those goals of improving communication among scientists and researchers.

From the cyberorganism perspective, it is only natural to investigate social phenomena such as knowledge growth, communication patterns, and social networks online. It is probably the first time that researchers in social sciences have an ideal research environment for conducting studies in social behaviours. Carrying out empirical research here combines the best characteristics from both the experimental control and the ecological validity standpoints. Here we have a public, controllable laboratory for conducting experiments which, by its design, gives us a perfect ecological validity; similarly, we have an ideal field ground for conducting observational studies where environmental variables are either under our control or at least predictable. In addition, online public behaviours can be tracked and documented through automated mechanisms, by utilizing the strange nature of the Internet—the cybernetic living system is completely artificial and public. By modeling its structure and processes, and how virtual cooperative interactions operate within it, we can systematically investigate naturally occurring phenomena on the net.

The following chapters will further examine of the cyberorganism. By the end of chapter 5 we will have a functioning model of the living Internet. In Part II of this dissertation, I will layout methodologies and techniques for conducting empirical investigations of virtual communities.

### *2.5 Summary*

The chapter traces the origins and evolution of the Internet. The first section explores its origins in cold war research environments, its actualization through development and growth, and its transformation into a global phenomenon. Memetic mixing and transmission via vehicles like *gametes* and *zygotes* are essential in its rapid development and growth. The second section describes the emergence of the global cyberorganism. It first depicts the world in which the cyberorganism inhabits and thrives: *cyberspace* is an objective, abstract, information environment delineated by Popper's *World 3*. The cybernetic, knowledge-creative, super-organism, called *cyberorganism*, is an aggregated living system based on *human-computer symbiosis*. Section 2 concludes with physiology of the cyberorganism. The last section examines the origins of *virtual cooperative interaction* and *knowledge creativity*. Scientific ethos, value systems, social norms are the deep roots for its motivation and reinforcement processes. The chapter concludes with suggestions for social sciences researchers that social behaviours on the net are ideal subjects for inquiries.

# CHAPTER 3

## Structures and Processes

A community is an open system. That is, a community maintains a relatively stable structure and boundary while receiving inputs from the environment, processing them, and extruding outputs. The human components of communities—individuals, groups and organizations—are also open systems. Furthermore, these open systems are composed primarily of living entities—cells, organs, and organisms (Tracy, 1989).

Viewed from this open systems perspective, the Internet community as a whole can be considered a living system just like a colony of social insects (such as honeybees) can be considered as a super-organism. Hence the hive is an organism and not merely the analogue to the person (Wilson, 1971). A super-organism like an ant colony defined by the eminent entomologist William Morton Wheeler (1911) has several important qualities that qualify it as an organism: (i) behaving as a unit; (ii) undergoing cycles of growth and reproduction that are clearly adaptive; and (iii) maintaining a complex, coordinated, living system. As we have seen in the last chapter about the origin, development and evolution of the net, the Internet can be construed as a global *cyberorganism* consisting of virtual organizations or special interest communities analogous to 'cyberorgans' which are in turn made up of 'cybercells'—people with network computers—the basic units of human-computer symbiosis.

Critical subsystems perform processes that are necessary for life and must be carried out by all living systems in order to survive or those processes must be performed for them by some other system (Miller, 1978). Typically, critical subsystems are not associated with any single component of the system and are not easily identified by their structure.

Rather, they are defined by the functional system requirements that they fulfill; that is, the vital processes they carry out.

Physiology and groupware system design focus on structures, whereas social psychology and human computer interaction concentrate on processes. Living system theory tries to keep both structures and processes in perspective, reminding us of the functional logic underlying structures.

## *3.1 Critical Subsystems in Living Systems*

Originally, the book *Living Systems* (Miller, 1978) presented 19 basic subsystems at seven levels, and since then, James Grier Miller and his long time collaborator Jessie L. Miller, have added a 20th subsystem, the *timer*, and an eighth level, the *community* (Miller & Miller, 1992). The community level is added between the organization and the society. The timer subsystem is added to the original list of nine information-processing subsystems, making a total of 10 (not counting the boundary and reproducer, which process both information and matter-energy). The 20 subsystems (shown in Table 1) are responsible for the ongoing day-to-day operation of the living system, it is these subsystems which keep the system alive (Bailey, 1994).

**Table 1 The Twenty Critical Subsystems of a Living System**

| **SUBSYSTEMS WHICH PROCESS BOTH MATTER-ENERGY AND INFORMATION** | |
|---|---|
| 1. Reproducer | |
| 2. Boundary | |
| **SUBSYSTEMS WHICH PROCESS MATTER-ENERGY** | **SUBSYSTEMS WHICH PROCESS INFORMATION** |
| 3. Ingestor | 11. Input Transducer |
| | 12. Internal Transducer |
| 4. Distributor | 13. Channel and Net |
| | 14. Timer |
| 5. Converter | 15. Decoder |
| 6. Producer | 16. Associator |
| 7. Matter-Energy Storage | 17. Memory |
| | 18. Decider |
| | 19. Encoder |
| 8. Extruder | 20. Output Transducer |
| 9. Motor | |
| 10. Supporter | |

*Source*: Adapted from Miller (1978); Tracy (1989) and Miller & Miller (1992).

### 3.1.1 Critical Subsystems

The survival and health of individuals, groups, organizations, communities, and societies depend on performance and coordination of a set of essential processes. In all, the living systems theory identifies 20 critical subsystems carrying out these processes at every level. For example, a channel and net subsystem to convey information from one part of the system to another was found to be necessary in cells as well as societies. The nature of that subsystem might vary but the function remained the same. Some critical subsystems process matter-energy, some process information, and some process both (Tracy, 1989).

A comprehensive presentation of the critical sub-systems in the living systems theory, is given by Miller (1978) and Tracy (1989). The following subsections provide brief descriptions and examples of the 20 critical subsystems in the living systems theory.

#### 3.1.1.1 Subsystems Processing Matter-Energy and Information

*Reproducer*. the subsystem which carries out the instruction in the genetic information (*template*) or character of a system and mobilizes matter, energy, and information to produce one or more similar systems. For example: in the group level, parents who create a new family; in the organization level, any individual, group, or department that produces a new organization with an implicit or explicit charter similar to that of the original organization; and at the community level, national legislature that grants state status to territory.

*Boundary*. The boundary is the subsystem at the perimeter of a system that holds together the components which make up the system, protects them from environmental stress, and excludes or permits entry to various sorts of matter-energy and information (Miller, 1978). At the group level, matter-energy: inspect soldiers of a platoon; information: TV-viewing rules in a family. At the organization level, matter-energy: security guards at entrance to a firm; information: librarian. At the community level, matter energy: agricultural inspection officers; information: movie censors in a town.

#### 3.1.1.2 Matter-Energy Processing Subsystems

*Ingestor*. The Ingestor is the subsystem that brings matter-energy into the system across its boundary. At the group level: refreshment chairperson of a social club. At the organization level: receptionists and personnel departments process inputs of people. And at the community level: airport authority of a city.

*Distributor*. The distributor subsystem carries matter-energy throughout the system wherever it is needed. The matter-energy may come from the environment through the Ingestor, from another internal process, or from storage. At the group level: father who serves dinner. At the organization level: assembly line. And at the community level: county school bus drivers.

*Converter*. The subsystem which changes certain inputs to the system into forms more useful for the special processes of that particular system. At the group level: work group members who cut cloth. At the organization level: operators of an oil refinery. And at the community level: city stockyard organization.

*Producer*. The producer subsystem takes matter-energy inputs directly or from the converter and synthesizes them into new materials. These new materials may be used by the system to provide energy, repair damage, or grow. They may also be extruded as trade goods or waste. At the group level: family member who cooks. At the organization level: machinery artifacts in manufactures. And at the community level: bakery and restaurant.

*Matter-Energy Storage*. The subsystem which places matter or energy at some location in the system, retains it over time, and retrieves it. At the group level: family members who put away groceries. At the organization level: stockroom personnel. And at the community level: county jail officials.

*Extruder*. The extruder subsystem transmits products and wastes across the boundary and out of the system. At the group level: kids who put out trash. At the organization level: employees in packaging, shipping, and mail room. And at the community level: city sanitation department.

*Motor*. The motor subsystem moves the system itself or parts of it, as well as components of the environment. At the group level: drivers of family cars. At the organization level: earth movers or cranes and their drivers of a manufacture. And at the community level: subway system and city transit authority.

*Supporter*. The support subsystem separates the various components of the system and maintains the proper spatial relationship between them so that they do not crowd each other. At the group level: housing construction crew. At the organization level: land and artifacts such as buildings and platforms. And at the community level: maintenance crew at the capital building.

### 3.1.1.3 Information Processing Subsystems

*Input Transducer*. The input transducer subsystem serves the same function with respect to information that the ingestor serves for matter energy. The input transducer brings information-bearing markers across the boundary and into the system. (A marker is a physical representation in the form of structure, pattern, or interaction of matter-energy.) At the group level: lookout of a gang of thieves. At the organization level: people in such departments as marketing research, sales, purchasing, legal, accounts and receivable, and product research. These people obtain information from the environment. And at the community level: representatives who report from a state capital to local voters.

*Internal transducer*. The internal transducer subsystem receives information-bearing markers from other subsystems or components of the system and converts the markers to nerve impulses. It is the sensory subsystem that receives markers changing them to other matter-energy forms of a sort which can be transmitted within it. At the group level: group member who reports members' opinions to group decider. At the organization level: employees who make internal reports on the status of, or changes in, variables of the system's components or subsystems, such as a factory quality control unit. And at the community level: neighborhood watch groups.

*Channel and Net*. The subsystem composed of a single route in physical space, or multiple interconnected routes over which markers bearing information are transmitted to all parts of the system. At the group level: person-to-person communication channels among group members. At the organization level: people at nodes of organizational networks, such as switchboard operators, secretaries, managers at all levels. And at the community level: telephone linesmen in a city.

*Timer*. The timer subsystem transmits information to a decider subsystem about time-related environmental states or about of components of the system. This information signals the deciders of subsystems to start, stop, alter the rate, or advance or delay the phase of one or more of the system's processes, to coordinate them in time. The timer consists of one or more oscillators known as clocks or pacemakers, the phase of which can be reset. They measure duration or order in time or underlying rhythms of various sorts. The timer synchronizes internal processes of the system and coordinates the system with its environment (Miller & Miller, 1992). For example, at the group level, a mother wakens other family members on time; at the organization level, workers take regular monthly inventory; and at the community level, artifacts such as clocks mark the opening and closing of schools and buildings, as well as regulating traffic lights and parking meters. The same can also be applied to a larger time scale, in terms of annual community celebrations of local and national holidays.

*Decoder*. Each living system has its private code for information. Data obtained through the input and internal transducers must be decoded and recorded into this private code by the decoder subsystem. At the group level: member who explains rules to a project team. At the organization level: manager who decodes the angry countenance of an employee into thoughts about the employee's attitude. And at the community level: attorney general of a state who interprets law.

*Associator*. The associator subsystem carries out all the learning processes of forming links or associations among various items of information. The associator is analogous to the producer in the sense that an association between two or more bits of information is a new bit of information. That is, putting bits of data together is like combining elements such as carbon and oxygen; the result is more than the sum of its parts. At the group level: parents who teach good behaviour to their children. At the organization level: people who train new employees. And at the community level: city school teachers, religious leaders in a neighborhood.

*Memory*. The memory subsystem completes the learning processes. It stores original bits of information and associations so that the total information in the system can grow over time. Memory processes involve input (recording or memorizing), maintenance (retention and recording as well as forgetting). and output (retrieval or remembering). Humans use artifacts like books, memos, diaries, files, and recordings to aid the memory subsystems. At the group level: father who keeps family records and history. At the organization level: accounting department that keeps financial records; filing department of a firm. And at the community level: operators of computer data bank at a central police department.

*Decider*. The decider subsystem is the most essential of all, according to Miller. It is the executive centre that receives data from various sources through out the channel and the net and sends control information (orders) to all parts of its system (Miller, 1978). Decision involves four stages: (i) establishing purposes and goals; (ii) analyzing discrepancies between the current state and variables and their desired state; (iii) synthesizing and choosing a plan of action to attain the desired state; (iv) implementing the plan by issuing appropriate commands (Tracy, 1989). Human artifacts of this subsystem include handbooks, algorithms, and computers. At the group level: parents; family council. At the organization level: the top echelon of the decider may be the board of directors as a group, or an individual such as the chief executive officer. And at the community level: governor, legislators, judges of a state.

*Encoder*. The subsystem which alters the code of information input to it from other information processing subsystems, from a “private” code which can be interpreted by other subsystems in its environment or by other living systems. It reverses the processes of the decoder. At the group level: writers of a group communication. At the organization level: billing, advertising, public relations, and legal departments that handle bills, advertising copy, and other messages that must be encoded for transmission to other systems. And at the community level: writers of city ordinances.

*Output Transducer*. The output transducer transfers information from internal markers to external markers that are suitable for carrying information in the system’s environment.

It changes the internal markers into other matter-energy forms which can be transmitted over the channels in the system's environment. At the group level: jury foreman who speaks for the jury as a whole. At the organization level: employees who deal with the public, such as salespeople, secretaries, spokespersons, and labor negotiators. And at the community level: representatives from a state to a regional legislature.

### 3.1.2 Relationships Among Subsystems

As stated earlier, one characteristic of life is that: all of the 20 subsystems are required for the survival and proper functioning of a living system, according to Miller. However a living system need not include all of these subsystems. It is possible to disperse certain functions to other systems. For example, a human fetus in the womb does not process food; the converter subsystem is supplied by the mother. A person may survive for at least a short period attached to a machine that supplies the distributor function of blood circulation. Motor functions lost because of a spinal cord injury may be partially recovered through various prosthetic devices (Tracy, 1989).

From the descriptions of the 20 critical subsystems as they manifest themselves in people and higher level living systems, it is obvious that there is much interaction among subsystems (Miller, 1978, Tracy, 1989). They frequently share components. A person's mouth is a component of the boundary, ingestor, extruder, input and output transducer subsystems. A networked workstation used by a group member in a software design team may be a component of its input transducer, decoder, internal transducer, decoder, channel and net, associator, memory, encoder and output transducer.

There must be a flow of matter-energy and information from one subsystem to another. Matter-energy is brought across the boundary by the ingestor and then carried by the distributor to the converter, producer or storage. If the flow is to the converter, from there the altered matter-energy may flow to the converter, producer, extruder, or motor. From storage matter-energy may flow to the converter, producer, extruder, or to any other subsystem to be used for growth and maintenance.

There is a similar flow of information through the boundary, input-transducer, channel and net, decoder, associator, memory, decider, encoder, and output transducer. Further more, there is substantial interaction between the matter-energy and information processing subsystems. Information is carried by markers of matter-energy; the information and its markers may be processed simultaneously. When a person eats a steak, the eyes, nostrils, and taste buds process the steak as information while the ingestor is processing it as matter-energy.

Proper functioning of all matter-energy subsystems depends on the flow of information to the decider, where control and coordination signals are issued to continue, stop, or alter the current processes (Tracy, 1989). Figure 7 provides a simplified picture of the relationships among critical subsystems.

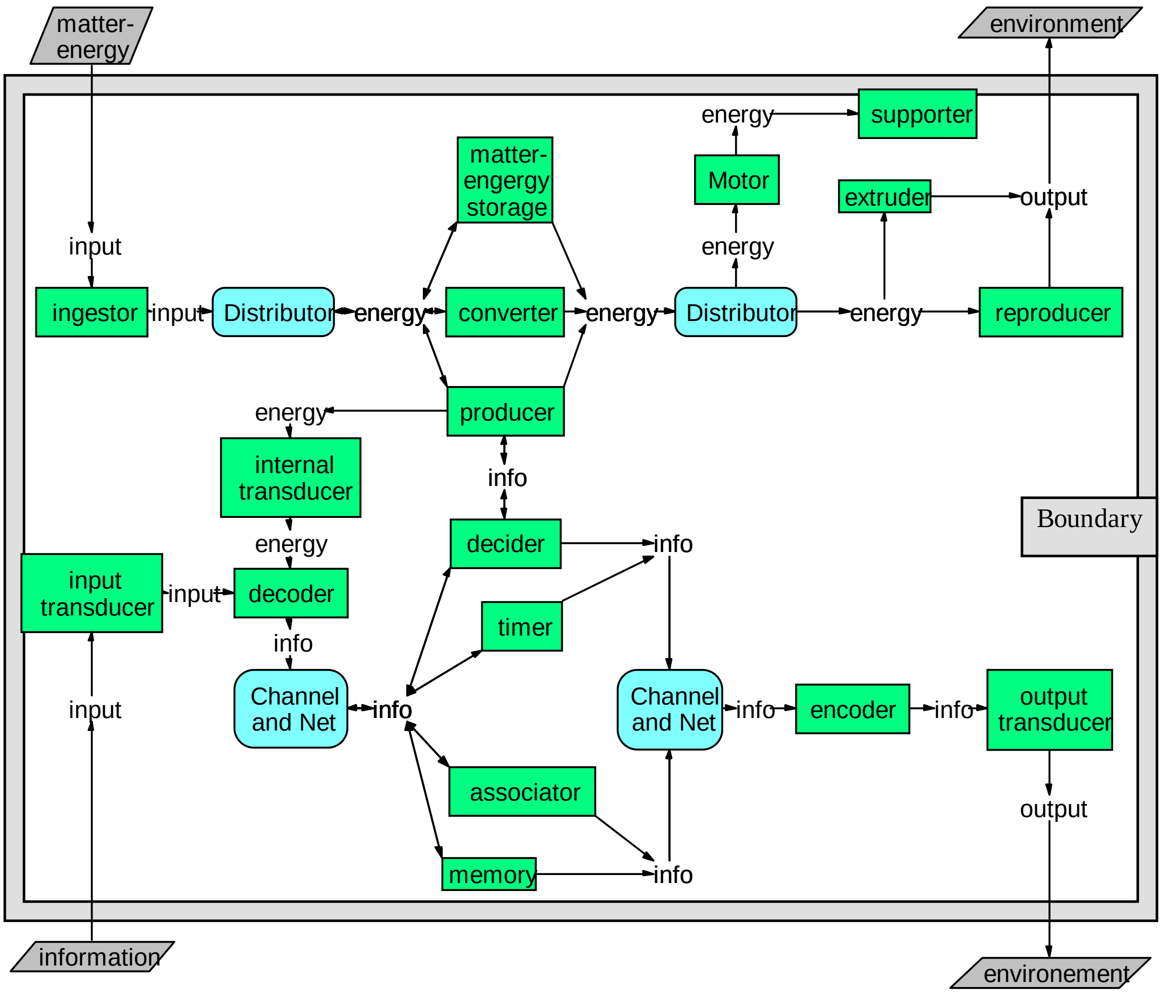


**Figure 7 Flow of Matter-Energy and Information**

## *3.2 Critical Subsystems in the CyberOrganism*

This subsection outlines some examples of sub-components which carry out the functionality of critical subsystems in the global cyberorganism. For example, computing networks (like the NSF backbone), networked workstations are the matter-energy infrastructures; and information infrastructures like the HTTP (hypertext markup language) protocols, web browsers, Alta Vista search engine are its information processing infrastructures (Chen & Gaines, 1997a).

*Reproducer*: Matter-Energy: corporations (usually multi-nationals) and governmental agencies (e.g., ARPA, NSF) that build computer networks and communication infrastructures which use the ARPANET or/and INTERNET as their blue-print/foundations. Information: Communication system-engineers who design new communication protocols such as HTTP 1.1, networking software, and new operating systems for the Internet and the World Wide Web.

*Boundary*: Matter-Energy: network routers, networked personal workstations; Information: computer networks administrators; referees of electronic scientific journals.

*Ingestor*: purchasing departments that are responsible for acquisition of computers and networking gears; local-telephone companies that act as common networks carriers; local power utilities which bring in electrical power source for the computer networks.

*Distributor*: long-distance telephone companies (e.g., MCI, AT&T, Merit) that provide the backbone infrastructures for the Internet.

*Converter*: packet switching network servers that converts raw data into data packets for transmission. e.g., ATM networks.

*Producer*: telecommunication and computer companies that produce workstations, networking hardware, etc.

*Matter-Energy Storage*: computing centres at university campuses that store computer equipment, spare parts, etc.

Extruder: network laser printers; data processing departments that produce billing letters, advertising mails, etc.

*Motor*: telephone linesmen; computer networks installation crews.

*Supporter*: chief information system officers in corporations who maintain over all integrity of company computing resources.

*Input Transducer*: data processing operators who input information; software programmers who input computer codes into computers; anyone who operates computers that are tied to the net directly or indirectly.

*Internal Transducer*: client-side browser programs like Mosaic and Netscape browsers that receives markers from the input transducers or other subsystems and convert them into HTTP or other protocols (e.g., FTP) for transmission in the net.

*Channel and Net*: communication technologies; gopher; World Wide Web; LANs; PAN (Personal Area Network).

*Timer*: chronological awareness support mechanisms like *CHRONO* or *WebWatch* that periodically update the information about the time-related states of the websites to people and/or other systems.

*Decoder*: operating systems that handle data packets and communication protocols like TCP/IP and HTTP.

*Associator*: search engines like Alta Vista; hyperlinks on web pages; subject hierarchy systems like *Yahoo*; editorial departments which keep *what's cool* site information; educational institutions that teaching computer skills; USENET newsgroups, MOO, IRC, and web conferencing systems (such as *HyperNews*) for collaborative learning between individual users or project groups that are often instigated by individual inquiry, followed by collective idea-generation process.

*Memory*: long-term memory: ftp archives, listservers, gopher and web documents. short-term working memory: USENET postings, e-mails (which could become long-term if necessary, however such short-term-memory to long-term-memory transformation are often selective).

*Decider*: net surfers; social networks like invisible colleges in virtual scholarly communities; industrial consortia; individual research groups and organization; technology policy departments in national government.

*Encoder*: CGI (common gateway interface) servers that encode relevant information like the results from knowledge databases into HTML for transmission to the web.

*Output Transducer*: artifacts including phone answering devices; fax machines; individuals or who get information from the Internet and act upon it, such as people who counsel the Internet information services for data retrieval and send information to others.

### *3.3 Summary*

In this chapter, 20 critical subsystems have been defined and their processes and structures have been described as they manifest themselves in human individuals and the cyberorganism. Structures provide a static picture of the arrangement of subsystems and components in a system, were processes have to do with dynamic changes in the matter-energy and information contained in the system. The primary structures of living systems are associated with the critical subsystems, each of which perform a function or set of vital processes for them. The cyberorganism framework has identified 20 such critical subsystems, governed by the decider subsystem. Although functionally distinct, these subsystems share components and interact closely. In next chapter, we will focus on key critical subsystems processing information involved in maintenance, control and coordination within the cyberorganism.

# CHAPTER 4

## Feedback, Control, Coordination, and Awareness

The primary imperative of life is immediate survival of the system through *maintenance of steady states*. In the near term a living system consists of a larger number of relationships or variable that must be held at or near steady states. Feedback processes within a living system are needed in maintaining steady states for performance (Miller, 1978). Within a complex living system, the subsystems themselves are also living systems. In addition, establishing feedback loops between living systems is essential in coordinating resource exchanges between them. In order for them to engage in mutual cooperative interaction, initially they need to be aware of the existence of one another. Afterward they need to be aware of what each is doing in order to continue the interaction. Here *awareness mechanisms* in a supra-living system transmit coordination signals among its subsystems or components. They constitute the feedback channels and regulate the adjustment processes within the supra-system.

For the purpose of this chapter, the decider, timer, net/channel, associator, and memory subsystems are most relevant in considering coordination processes in the cyberorganism. The decider subsystem is the most essential subsystem, according to Miller. It is the executive centre that receives data from various sources through out the channel and the net and sends control information to all parts of its system. Awareness can be viewed to be coordination signals sending among subsystems. Finally, this chapter examines the net/channel, associator, and memory systems in the cyberorganism. These subsystems' functional activities correspond to the communication, social and knowledge processes frequently observed on the Internet

### *4.1 Flow in Feedback*

Those processes within living systems which maintain steady states are *adjustment processes*. They operate based on the notion of feedback. When signals are fed back over the feedback channel in such a manner they increase the deviation of the output from a steady state, *positive feedback* exists. When the signals are reversed, so they decrease the deviation of the output from a steady state, it is *negative feedback*. Positive feedback alters variables and destroys their steady states in systems. It can initiate system changes and growth. Negative feedback maintains steady states in systems. It cancels an initial deviation or error in performance. Careful maintenance between positive and negative feedback processes are important criteria in balancing the two imperatives of life: actualization of the system's potential through the development of complexity and the maintenance of steady states (Miller, 1978). When such balance in feedback processes is achieved, a sentient living system is engaging in its optimal experience called **Flow**—the process of optimal experience—it is achieved "*when a sufficiently motivated user perceived a balance between his or their skills and the challenges of the interaction, together with focused attention*" (Csikszentmihalyi, 1990).

In terms of the fundamental level of human-computer symbiosis involving the net, Hoffman and Novak (1995) have applied the model of *flow* phenomena in user interaction with computer-mediated hypermedia, such as the web.

One possible explanation for the growing popularity of *net surfing* may be that such playful activity provides users a great sense of satisfaction. It is a state in which a motivated user undertakes a task whose level of difficulty is at some particular level that suits their individual needs. Too low a level results in boredom and too high a level in anxiety, and the optimal level results in the intense satisfaction with the *flow activity*.

Network navigation on the web makes optimal experience easier to achieve, because it has rules that require the learning skills, developing goals, providing feedback, and making control possible. Flow formalizes and extends a sense of playfulness, incorporating the extent to which, in the hypermedia environment, web users: (i) perceive a sense of *control* over their interactions in the environment, (ii) focus their *attention* on the interaction, and (iii) find it *cognitively* enjoying. When in the flow state, irrelevant thoughts and perceptions are screened out and the users' attention is focused entirely on the interaction (Hoffman & Novak, 1995).

Flow thus involves a merging of actions and awareness, with concentration so intense there is little attention left over to consider anything else. A net user's action in the flow state is experienced "*as a unified flowing from one moment to the next, in which he is in control of his actions, and in which there is little distinction between self and environment, between stimulus and response, or between past, present, and future*" (Csikszentmihalyi 1990; Csikszentmihalyi, 1993). Self-consciousness disappears, the user's sense of time becomes distorted, and the resulting state of mind is extremely gratifying.

## *4.2 Decider Subsystem*

The decider subsystem is the most essential of all, according to Miller. It is the executive centre that receives data from various sources through out the channel and the net and sends control information (orders) to all parts of its system (Miller, 1978). Since the net is fundamentally a distributed system, we need to take into account how the decision controls can be distributed among its constituent components or subsystems.

The *Collective intelligence model* proposed by John B. Smith (1994) can be characterized as a cognitive model for the decider subsystem in a cyberorganism. It focuses on two system levels: the individual and the group levels (the lightly shaded inner areas in Figure 3). The model regards collaborative groups as a form of information processing system, analogous to Newell and Simon's Information Processing System (IPS) model of individual cognition (Newell, 1982; Newell & Simon, 1972).

A collective processor includes the fine-grain operations used by groups to develop, access, and maintain the information stored in the memory system. The collective processor as a whole can be viewed as a loosely coupled distributed system that includes multiple independent processors, joined by communications and social networks.

Within the model, a collective memory system includes subsystems that provide a collective long-term memory for tangible knowledge, built and maintained in a computer system, and for intangible knowledge, carried in the heads of the human beings that comprise the group. The memory system also includes working memory (the currently attenuated awareness) for both types of information (Smith, 1994).

Collective strategy enables coherence in collaborative work. Individual processes occur not in isolation but in purposeful sequences. These strings of operations are analogous to statements in a language intended to accomplish a goal or to communicate a message. The system responsible for generating sequences of operations is analogous to the grammar individuals used to generate a string of words. Therefore, a collection of collective strategies become a work-flow model for the collaborative group.

Finally, the model considers two meta-cognitive issues: collective awareness and collective control. Many collaborative projects are too large and too complex to be understood by any one person. However, people often expect groups to produce work with the same integrity and consistency sometimes found in work produced by a single good mind working alone. By developing thick, overlapping areas of shared knowledge, groups may be able to piece together a form of collective, but distributed peripheral awareness that is sufficiently coherent to achieve this goal. Control must also be distributed over a group. Otherwise, information will not flow across boundaries, and the group and its work will be brittle. Here *awareness maintenance* provides essential coordination signals for groups to exercise distributed, collective controls that maintain smooth coordination (Chen & Gaines, 1997a). However, although many decisions can, and probably be made by consensus, authority must ultimately be centralized in order to resolve disagreements and to preserve the integrity of the group’s work.

### *4.3 Awareness Maintenance and Computer-Supported Cooperative Work*

A critical requirement in shared tasks is maintaining *situational awareness* (Norman, 1993) by keeping everyone adequately informed. In an environment where each member has a well defined role, the need to have face-to-face communication in order to perform a cooperative task becomes less necessary if mechanisms for situational awareness have been well established between members. For example, the navigation of a large ship requires effective coordination of various people with differing roles (Hutchins, 1990). Many key members of the navigation team (conning officer, plotter, bearing takers, deck log keeper, bearing-timer recorder and fathometer operator) are geographically separated (pilot house, chart house, and port and starboard wings) and communicate with each other by a common telephone circuit. The common audio channel and the physical layout of the pilot house provide opportunities for the navigation crews to observe each other's work, contributing to partial redundancy in their joint knowledge. They also support maintenance of the group over time to provide fault-tolerance if some group members fail to perform their roles.

Thus, one of the important criteria for achieving group cohesiveness is not the opportunity of being in constant face-to-face communication, but rather the situational awareness of what other group members are doing. It can be achieved through either face-to-face (as on the pilot house) or telephone communication (as between the port and starboard wings). Together the functional specificity of the crews and the cognitive artifacts that facilitate situational awareness (e.g., the single telephone circuit, the pilot house with high visibility among navigational team members) can create an effective collaborative system.

The emphasis on the importance of social interaction and cognitive artifacts (such as the telephone circuit described previously or the web) as the means to enhance human capabilities has also been echoed by Norman (1991). A *cognitive artifact* is defined as "an artificial device designed to maintain, display, or operate upon information in order to serve a representational function." The power of a cognitive artifact comes from its function as a representational device. However, capabilities of artifacts do not actually change an individual's abilities. Rather, they change the nature of the task performed by the person. When the informational and processing structure of the artifact is combined with the task and the informational and processing structure of the human, the result is to expand and enhance cognitive capabilities of the total system of human, artifact, and task.

Hence, when designing a computer-supported collaborative system, designers need to keep the representational aspect of the whole system in mind. A system designed with this holistic perspective in mind can provide tools that enhance communication, coordination and social interaction capabilities. Therefore, the collaborator's intentions can be easily transferred to each other and common goals can be achieved with mutual satisfaction. In the various studies described above, situational awareness has acted as an important coordination mechanism among collaborative groups.

Research in computer-supported cooperative work (CSCW) focuses on assisting people to work collaboratively as a cohesive team and providing them with a sense of common purposes (e.g., completion of the group task). For example, Landow's (1990) *In Memoriam* project utilizes hypertext's freedom of navigation and linking ability to break down physical separation and the univocal voice of textual conversation. In so doing it creates a new awareness of the processes of collaborative learning and collaborative work for group members in literary studies. Ishii and colleagues (Ishii and Miyake, 1991; Ishii and Kobayashi, 1992) have created a fusion of video and computer workspace to provide a seamless working environment with remote workstations, in order to provide mutual awareness between distance collaborators in real-time. Olson and Atkins' (1990) *NSF EXPRES* Project uses intelligent, multimedia email to facilitate cooperation within scientific and engineering community by increasing researchers' awareness about each other's work. Gutwin and Greenberg's (1995) group awareness widgets for desktop conference provide basic tool-kits to construct groupware that help tele-conference participants to be aware of: each other's focus when their views are separated; others' task activities in shared and separate view situations; and the history of group activities.

One application of those awareness widgets is to providing “workspace awareness” in collaborative learning, i.e., the up-to-the-minute knowledge of a learner requires about other learners’ interactions with the shared workspace (Gutwin, Stark, & Greenberg, 1995).

### *4.4 Awareness in Collective Intelligence*

In a dynamic environment where large amounts of information are created and updated frequently, the need to keep up with the most up-to-date and relevant information has become more important as the Internet community expands. In terms of group collaboration, to be aware of changes is one of the fundamental requirements for coordination and for providing a sense of connectedness.

Some of the characteristics of human intellectual work that are valued most highly are: coherence, consistency, correctness, and elegance. It is difficult to imagine how work with these attributes could be produced without that structure of ideas having been held in its entirety by a single mind—if not actually produced by that mind (Smith, 1994). As stated previously, however, by considering awareness from a functional point of view, one may be able to construct mechanisms that enable groups to achieve comparable result. Web pages (in addition to FTP archives, listservers, etc.) are becoming the primary means for information dissemination on the Internet, and these web pages are being constantly updated to reflect members’ current states of knowledge on their portions of collective memory.

The following subsections are detailed descriptions of a collective awareness taxonomy based on the *collective intelligence* model. Using the human self-awareness model and cognitive IPS architecture (Newell and Simon, 1972; Newell, 1982) as references, two analogous forms of awareness can be identified for collective groups: awareness of the group’s collective long-term memory and awareness among and of each other.

### 4.4.1 Awareness of the Group's Collective Long Term Memory

The collective long-term-memory has two parts: the artifact (the objective of the group task) and the body of shared intangible knowledge (that is, the information carried out in the heads of group members originally). Since awareness of the artifact exists only in the minds of the human beings who comprise the group, awareness of the artifact can also be seen as part of the group's intangible knowledge. Awareness in groups exists at several levels of detail:-

*Global Awareness* —The most general level is the body of intangible knowledge that is shared by all members of the group. It includes the overall goals of the project, its ways of operating, the strategies it uses to develop the artifact, its current status and problems, the relation of the project to the external environment, etc. This awareness is not deep, if the project is large, but it provides each member with a sense of the whole.

*Deep Awareness* —At the other extreme is the deep, detailed, often technical, knowledge held by individual members. Depending on the project, a single individual is often responsible for a particular part. Thus, the level of awareness and expertise required to generate a segment of the overall artifact (project) is significantly greater than that required for another person to understand it.

*Peripheral Awareness* —Between the extreme of general, shared knowledge and deep, individually specialized and generative knowledge, is an intermediate level: the thick knowledge of adjacent or near specialty areas. It takes the form of understanding, rather than generation. Thus, it is shared with individuals or the team responsible for developing other parts of the artifact, but it is not as deep as their knowledge nor is it shared with the entire project. This critical peripheral awareness is ultimately responsible for the integrity of the group's work. It provides a context for the interfaces between areas. Thick, shared knowledge can be developed through informal interactions, such as conversations, but it can also be developed through more formal mechanisms, such as institutionalized reviews (e.g., structural walk through process in a software project).

Thus, the model identifies three forms of awareness with respect to a group's long-term-memory:-

1. close, detailed, *deep awareness* of particular segment of the artifact;
2. less detailed, but still substantial, *peripheral awareness* of the artifact's adjacent parts;
3. the much thinner, *global awareness* of the artifact as a whole that is shared by the entire group.

### 4.4.2 Awareness Members have of One Another

A different kind of awareness is the awareness members have of one another. This category of awareness is closer to the notion of *situational awareness* described earlier.

*Resource Awareness* —One of the primary reasons for assembling a group is to assemble the expertise required to carry out a project. Therefore, the issue becomes one of providing the group (as a whole) with a collective awareness of its members' respective specialized knowledge and expertise. Thus, an extremely valuable resource for a group is shared knowledge of who is an expert on what.

*Task-Socio Awareness* —Another form of awareness involves the interaction between social and intellectual processes operating within the group. It would be simple if groups were purely intellectual organisms, but they are not: tensions exist and factions develop. These developments are inevitable. For example, one member may oppose an idea voiced by another, not because the idea is objectionable but because of who said it. The opposite condition —support an idea because of friendship or attraction —is equally bad. These so very human situations are unlikely to go away, but a group should be aware of itself as a dynamic, functioning organism as well as be aware of the artifact it is developing to insure that the integrity of its work is not compromised by them.

*Chronological Awareness* —A third form of awareness is the instantaneous awareness that an individual has regarding the activities of other individuals. This is what constitutes chronological awareness. For example, one member of the group may be aware (or wish to know) what another member of the group is working on in a nearby part of the artifact. This behavior is monitored at a very low level by the collaboration support system in its concurrency control mechanisms to insure that two members do not attempt to change the same part of the artifact at the same time. These mechanisms, however, do not prevent one member's access from blocking that of another, or prevent one member's subsequent work from affecting earlier work done by another. Groups may need help in monitoring domains of activity. For example, members may want to see where colleagues are working; they may even wish to see a display over time of the "tracks" left by colleagues.

Thus, the model identifies three forms of awareness among members of one another:-

1. cognitive-guide-map like, *resource awareness* for locating specific knowledge and expertise among the group members;
2. emotional, but rationalized, *task-socio awareness* concerning social and political dynamics within the group in relation to the artifact/task.
3. and more situational, *chronological awareness* about when and what by whom something in the collective system is changing.

The above *collective awareness* taxonomy provides a useful conceptual framework for investigating awareness mechanisms in the cyberorganism. Those corresponding awareness services can be seen in right branch in the general taxonomy of net services (Figure 5).

### *4.5 Net/Channel, Associator, and Memory Subsystems*

This section examines the net/channel, associator, and memory systems in the cyberorganism. These subsystems' functional activities correspond to the communication, social and knowledge processes frequently observed on the Internet. First the net/channel subsystems are comprised of: discourse and file transfer mechanisms (for communication processes). Second, the associator systems are special interest communities (for social and knowledge creation/assimilation processes). Third, the memory subsystems are resource storage mechanisms (for knowledge discovery and retrieval processes).

Figure 8 shows a typical configuration to establish when setting up integrated support for associator subsystems—special interest communities on the net—where social and learning processes occur. The net and channel subsystems include: e-mail, www, and ftp mechanisms. A mailing list server handles the primary discourse of the communities. If there are distinctive sub-groups or activities it may be appropriate for it to manage several mailing lists for a single community. A file transfer protocol (FTP) server handles access to the document archives, supporting both uploading and downloading arbitrary files. A hypertext transfer protocol (HTTP) server handles access to HTML documents through World-Wide Web (Gaines, 1994a).

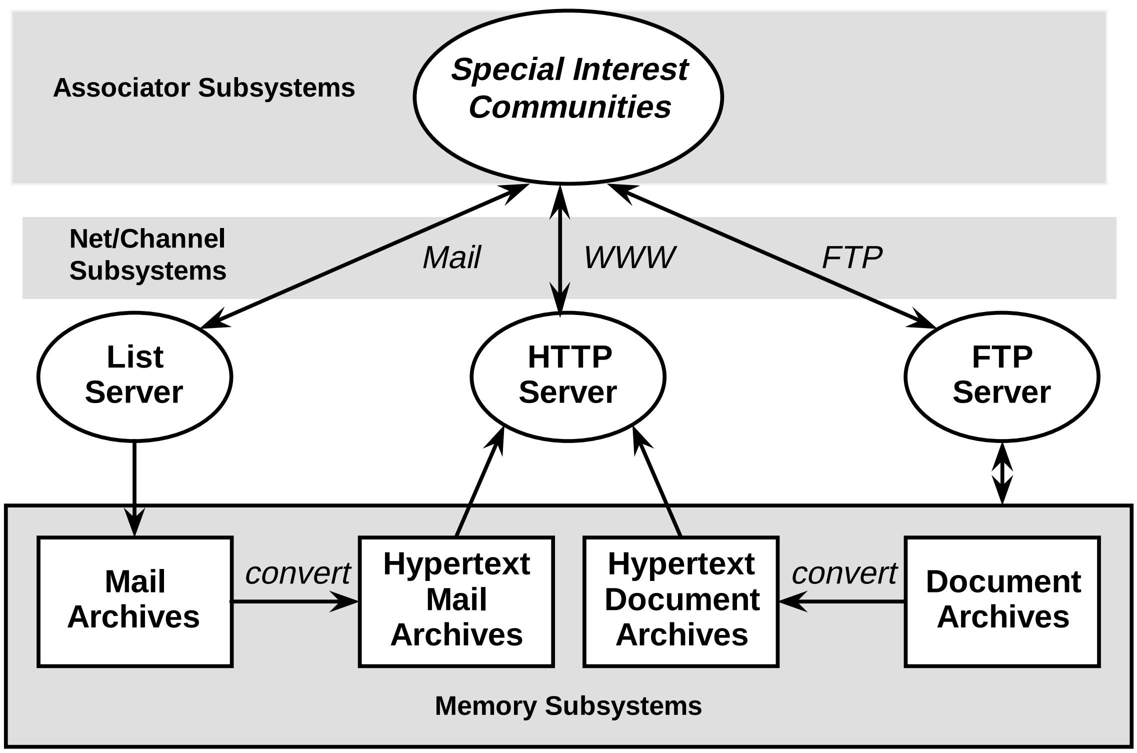


**Figure 8 Integration of Net/Channel, Associator and Memory Subsystems**

The memory subsystems include: archival mechanisms. The mail archives are made available through FTP, and they are also automatically converted to HTML format to allow them to be browsed easily through the web. For example, the archives of the www-talk list server has been converted to HTML and indexed by topic using the conversion program *Hypermail*. The FTP server gives access to all the archives, allowing new HTML documents to be uploaded to maintain the web facilities.

### *4.6 Summary*

Feedback processes within a living system are needed in maintaining steady states of performance which is the primary imperative of life. *Awareness mechanisms* in a supra-living system transmit coordination signals among its subsystems or components. They constitute the feedback channels and regulate the adjustment processes within the supra-system.

This chapter investigates the decider, timer, net/channel, associator, and memory subsystems in respect to control, coordination and awareness processes in the cyberorganism. The decider subsystem is the most essential subsystem, according to Miller. It is the executive centre that receives data from various sources through out the channel and the net and sends control information to all parts of its system. Awareness can be viewed to be coordination signals sending among subsystems. The chapter surveys concepts and models in the CSCW research with respect to coordination and awareness processes. Collective intelligence and collective awareness can be considered cognitive models for the coordination purposes of the decider, timer, associator, and memory subsystems. Finally this chapter examines an integrative architecture for the net/channel, associator, and memory systems in the cyberorganism.

The next chapter presents three aspects of virtual cooperative interaction within the global cyberorganism (i.e., the Internet community as a whole). The main emphasis is trying to address the communication, social and knowledge processes within a cyberorganism. It consists of three aspects: (i) the descriptive aspect which characterizes and classifies virtual cooperative interactions; (ii) the prescriptive aspect that provides motivational reasons for individuals to participate in virtual cooperative interactions; and (iii) the operational aspect of how virtual cooperative interactions initiate and function.

# CHAPTER 5

# Elements of Virtual Cooperative Interaction

The last chapter introduced feedback, control, coordination and awareness processes in the global cyberorganism. It surveyed the research in CSCW which involves many aspects of awareness for the coordination of group tasks and interactions. Using CSCW terminology, the cyberorganism can be considered as a very-large-scale *groupware* system, which supports complex cooperative interactions among variety of net users. Those cooperative interactions differ qualitatively from goal directed, highly coordinated team-oriented collaborations supported by conventional groupware. The emphasis in this chapter is on virtual cooperative interaction.

## *5.1 CSCW Beyond Groupware*

In large-scale organizations, geographically separated sub-units may be working together in joint collaborations. With geographical separation, it is difficult for each unit to keep the other continually in mind and, therefore, to keep their common goals in mind. Naturally, most people attend to what is close at hand and ignore what isn't. In addition, priority isn't perfectly correlated with proximity, hence remote communications are essential even in highly decentralized organizations. Yet they can be difficult (Kiesler & Sproull, 1991).

Increasingly there are commercial pressures to decentralize and a growing recognition that organizations can successfully conduct their business by providing a distributed workforce with cost effective telecommunications devices. The expectation is that cost-effective telecommunications can recreate a fully functioning virtual organization. (Sheehy & Gallagher, 1996).

Many organizations are just starting to incorporate computer-mediated communication technology to create new inter-organizational linkages, to solve out-of-sight, out-of-mind problems, and to create dynamic structures (Kiesler & Sproull, 1991). Increasing the communications facilities available at the networked workstation affords new opportunities for organizational members to sustain a high level of routine conversational flow (Sheehy & Gallagher, 1996). Creating global virtual organizations through the net will inevitably alter communications processes that create a sense of what it feels like to be with an organization.

In addition, diffuse collaborative communities that have formed on the Internet suggest that there is a need to reconsider existing notions of CSCW. Large scale groupware differs not only in the quantity, but also in the quality of cooperative interaction. The majority of CSCW research has focused on collaborative interactions between members of a task-oriented, purposeful work-group. For example, Ellis and Wainer (1994) presented an ontological/coordination/user-interface model for task-oriented groupware; Dourish and Bellotti (1992) discussed awareness and coordination issues in shared workspaces of co-authoring systems.

Recently some CSCW researchers have started to investigate cooperative interactions beyond cohesive, task-oriented small-group collaboration. For example, research has been conducted on purposeful large-scale community-wide collaborative systems (Star & Perrochon, 1994); on informal large-scale community-wide messaging systems (Brothers, Hollan, Nielsen & Stornetta, 1992; Resnick, Iacovou, Suchak & Bergstrom 1994); on design principles for online communities (Kollock, 1997; Mynatt, Adler, Ito & O'Day, 1997); and on social filtering systems (Hill, Stead, Rosenstein & Furnas, 1995; Hill & Terveen, 1996).

## *5.2 Virtual Cooperative Interaction within the CyberOrganism*

As stated earlier in Chapter 1, the fundamental nature of interaction on the net/web can be characterized as *virtual cooperative interaction*. The word "virtual" has two senses here:

first, it denotes the notion of *virtual space*, i.e., the cooperative interaction that occurs in a non-physical space which allows participants to be situated in geographically separate locations; second, it denotes that the *intention* to engage in cooperative interaction itself may not necessarily pre-exist or be conscious. Traditional CSCW research focuses on the first sense (tele-presence in virtual space), but there is a need to extend the notion of cooperative interaction to encompass the latter sense of virtual cooperative interaction also (Chen & Gaines, 1997b).

The main purpose of *cooperative interaction* is to propagate resources and memetic transmission among living systems. The *systems dyad* (shown in Figure 2) emphasizes the fact that the typical exchange between two systems is *mutualistic* (Dunbar, 1988), and that other originator systems and receiver systems may exist. In some cases the interdependence between the systems becomes so great the dyad itself becomes a supra-living system. *Virtual cooperative interactions* are critical coordination processes for establishing mutualistic relationships among subsystems within a supra-living system.

Virtual cooperative interaction is involved with the exchange of resources between subsystems or components in a cybernetic living system. Establishing feedback loops between subsystems is essential in coordinating resource exchanges between them (Miller, 1978). Therefore, communication processes through the channel and net critical subsystems. In order for them to engage in mutual cooperative interaction, initially they need to be aware of the existence of each another. Afterward they need to be aware of what each other is doing in order to continue the interaction. Here, *awareness mechanisms* in a supra-living system transmit coordination signals among its subsystems or components. They constitute the feedback channels and regulate the adjustment processes within the supra-system. Human living systems are the fundamental vehicles for memes and critical subsystems for carrying out cooperative behaviour processes in virtual cooperative interaction. Discourse processes and communication structures formulate the net and critical channel subsystems. Psychological motivations and social reinforcements are crucial in initiation and continuation of virtual cooperative interaction.

Frequently, information resource contribution and exchange on the web involve cooperative interaction without pre-planned coordination. In fact, participants on the web may have no intention to cooperate in the first place. Quite often, a resource provider and a resource user are unaware of each other's existence until their first interaction. Nevertheless, the interactive process between them is still loosely cooperative in nature. It differs from the traditional team-oriented cooperation where group tasks, goals, and purposes are usually well-defined.

A classical *social exchange* model like the Interactional Matrix model (Kelly & Thibaut, 1978; Cook, 1987) cannot readily account for this unusual form of cooperation where a resource provider might never know the identity of her resource users, and yet still continues to contribute anyway. On the web the only feedback she may receive might be the frequency of accesses to her information resources. What does she gain in return in such a seemingly one-way cooperative interaction? Is it simply an expression of altruism? What are some possible motivations for her to contribute to the web? In general, how would one ensure the continual contribution of an information provider? These questions can be answered more clearly in the context of *socioware*.

The present chapter defines **socioware** as: *subsystems or components within a cyberorganism for supporting communication, knowledge, and social processes which expedite virtual cooperative interaction*. Socioware emphasizes the second definition of virtual cooperative interaction as contrast to groupware. Information inquiry and response, dissemination of ideas, and social networking are some examples of virtual cooperative interaction. USENET newsgroups and listservers are two prototypical socioware that support dialogues within well defined special interest communities on the net.

The proliferation of personal home pages with cross-linkage of web pages by people who share common interests has made the exploration process on the web (i.e., net surfing) a potentially *social experience*. Such a seemingly intrinsic rewarding experience can often be characterized as serendipitous and not necessarily task-oriented (as in traditional groupware). Through home pages, individuals create their own *virtual persona* on the web without any awareness of whom their eventual audience might actually be (i.e. without *extensional awareness* of particular recipients). However, they often have a sense of who the potential audience might be (i.e. with *intensional awareness* of the type of recipient). Sometimes individuals provide an information resource to the web as a by-product during some self organization processes of their own knowledge. As observed earlier, this form of apparently cooperative behavior is prevalent on the web (Chen & Gaines, 1996a).

In essence, the goal of socioware is to facilitate emergent pro-social cooperative behaviours for self-organized, virtual communities within the cyberorganism.

This chapter describes a socio-psychological model that encompasses collaborative activities supported by traditional groupware and by socioware within the cyberorganism. The model analyzes the following five basic elements involved in virtual cooperative interaction:

1. *discourse patterns*
2. *time-dimension of virtual interactions*
3. *awareness hierarchy*
4. *motivations for cooperative behaviors*
5. *emergence and maintenance of virtual cooperative interaction*

Together they present three aspects (what, why, and how) of the model: (i) the descriptive aspect comprised of the first three elements which characterize and classify virtual cooperative interactions; (ii) the prescriptive aspect that provides motivational reasons for individuals to participate in virtual cooperative interaction; and (iii) the operational aspect of how virtual cooperative interaction initiates and continues.

## *5.3 Some Definitions*

Before describing the conceptual model in detail, the definitions of some frequently used terms in the model are introduced in this section.

The term **social entrainment** refers to some endogenous biological and behavioral processes that are captured, and modified in their phase and periodicity, by powerful (internal or external) cycles or pacer signals emitted by the timer critical subsystem. The notion of entrainment contains two kind of synchrony: (i) The *mutual entrainment* of endogenous rhythms to one another; (ii) the *external entrainment* of such a rhythm by powerful external *signals* or *pacers* (McGrath, 1990).

When individuals participate in virtual cooperative interactions, depending on the nature of their present focus (e.g., discuss an idea, co-reviewing a book), there is a natural cognitive processing time involved in each activity. This processing time generates an endogenous rhythm within individual participants. This natural rhythm of interactions consequently creates *mutual entrainment* in sustaining continuation of virtual cooperative interactions. The processes of social entrainment are important in the time-dimension of virtual cooperative interaction. They essentially embody the main purpose of the *timer* subsystems within living systems.

One can regard a collaborative community as a set of individuals that provide resources to one other with the most significant dimension relating to the coordination of the community being that of the **resource awareness** of who is providing a particular resource and who is using it (Chen & Gaines, 1997a). Logically, resource awareness can be further distinguished as *extensional awareness* because the specific resource and provider are known. It is contrasted to *intensional awareness* in which only the characteristics of suitable resources or providers are known.

## *5.4 Systems Dyad for Punctuated Discourse*

At first glance the recent development on the net has been evolving away from conversation, toward demonstration. This general trend is typified by the recent appearance of the web, splashy multimedia uses of the Internet, and the virtual explosion of voice, color, picture and motion seemed to overshadow plain ASCII traffic in ideas and emotions (Rafaeli, McLaughlin & Sudweeks, 1997). Nevertheless the essence of the net remains categorically a medium for conversation (McLaughlin, 1984). Interactivity is one of the key qualities of discourse made possible by computer-mediated communications (Rafaeli, 1988; Rafaeli & Sudweeks, 1997). It remains a defining feature for conversations generally, and an especially curious one when the conversations are not held in physical proximity (Rafaeli, McLaughlin & Sudweeks, 1997). This section proposes an integrative model of discourse patterns that incorporates the current major innovations and services on the Internet.

Figure 5 in chapter 2 presents a conventional taxonomy of Internet services in terms of their utility. And Figure 8 in the last chapter presents an integrated architecture consisted of channel/net, memory and associator critical subsystems. Individuals represent the fundamental cognitive *agents* within the associator subsystems—i.e., special interest communities in the cyberorganism.

However, the taxonomy and architecture do not convey the dynamic processes of how the cyberorganism supports virtual communities. A dynamic model can be developed by noting what distinguishes discourse from publication is that in discourse it is expected that the recipient responds to the originator, whereas publication is generally a one-way communication. However, on listservers some material is published in that the originator expects no specific response, and material published in electronic journals or archives often evokes a response. One useful perspective for studying group computer-mediated communication is interactivity. As noted previously (Rafaeli, 1988), there are at least three modes of communications on the net: one-way, two-way (reactive), and interactive (Rafaeli & Sudweeks, 1997). Hence, computer-mediated communication offers a very flexible medium that breaks down the conventions of other media. The following diagrams show the different characteristics of the main Internet services in terms of these issues.

Figure 9 shows email discourse as a cycle of origination and response between a pair of agents communicating through a computer-mediated channel.

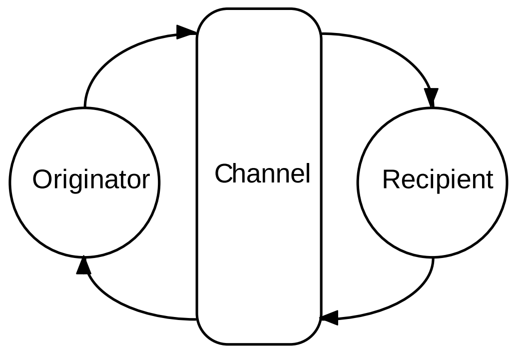


**Figure 9 Email Discourse**

Figure 10 extends Figure 9 to show listserver discourse as a cycle of origination and response between agents that is shared with a virtual community through a computer-mediated channel. The virtual cooperative involvement leads to more complex discourse patterns in that: the originator may not direct the message to a particular recipient; there may be multiple responses to a message; and the response from the recipient may itself trigger responses from others who did not originate the discourse. For a particular discourse sequence, this leads to a natural division of the virtual community into active participants who respond and passive participants who do not.

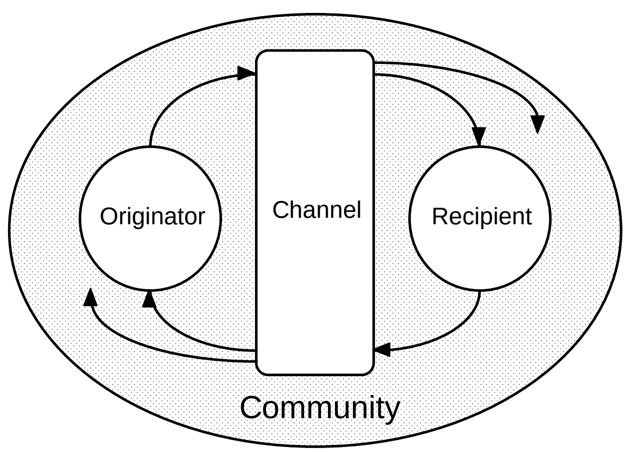


**Figure 10 Listserver Discourse**

Figure 11 modifies Figure 10 to show web publication as an activity in which the channel is buffered to act as a memory store also. The material published is available to a virtual community and the originator is unlikely to target it on a particular recipient. Recipients are not expected to respond directly to the originator, but responses may occur through email, listservers or through the publication of material linked to the original. Because the published material is not automatically distributed to a list, recipients have to actively search for and discover the material.

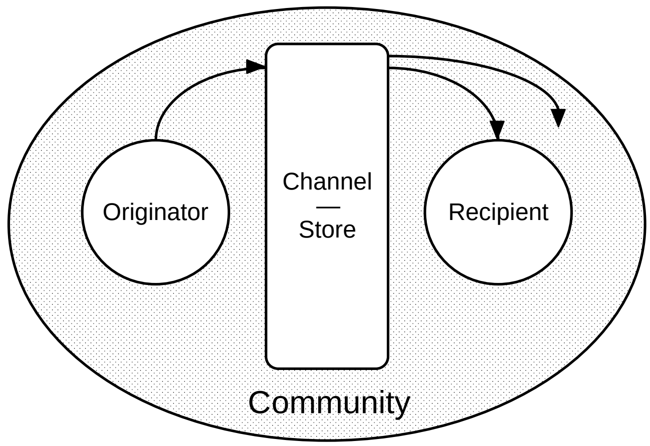


**Figure 11 World Wide Web Publication**

The common structure adopted for the diagrams is intended to draw attention to the commonalties between the services. Listserver discourse is usually archived and often converted to hyper-mail on the web. Web publications do trigger responses through other services or through links on the web. A search on the web may not discover a specific item but rather a related item on a newsgroup, list or by an author, and result in an request for information to the newsgroup, list or author. Individuals and communities use many of the available Internet services in an integrated way to support their knowledge processes.

Figure 12 subsumes Figure 9 through Figure 11 to provide an integrated model of communication, knowledge and social processes in cyberorganism. It captures all the issues discussed. It models the processes as discourse punctuated by the intervention of a memory store allowing an indefinite time delay between the emission of a message and its receipt. It introduces two major dimensions of analysis: the *time* for each step in a discourse cycle; and the *awareness* by originators of recipients and vice versa.

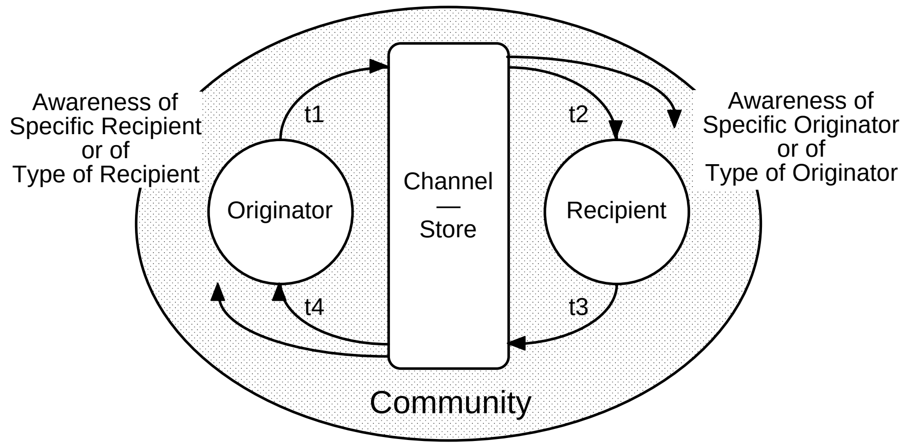


**Figure 12 Punctuated Discourse**

## *5.5 Time Structure of Punctuated Discourse*

The four times shown in Figure 12 are:

| | | |
|---|---|---|
| **t1:** | *origination time* | the time from a concept to its expression and availability |
| **t2:** | *discovery time* | the time from availability to receipt |
| **t3:** | *response time* | the time from receipt to expression and availability of a response |
| **t4:** | *response discovery time* | the time from response availability to receipt |

Note that agent processing times and channel delays have been lumped. A study focusing on the impact of communication delays would want to consider them separately; otherwise there is no significant distinction—a general principle might be that communication delays should not be greater than agent processing times. Note, also, that the diagram is to a large extent symmetrical—the recipient becomes an originator when responding.

An important overall parameter is **time cycle**: the round-trip discourse time, t1+t2+t3+t4. If this is small, a few seconds or less, we talk in terms of synchronous communication. If it is large, a few hours or more, we talk in terms of asynchronous communication. If it is infinite, so that there is no response, we talk in terms of publication. However, this analysis shows that there is a continuous spectrum from synchronous through asynchronous to publication.

The discovery times, t2 and t4, are very significant to publication-mode discourse, and attempts to reduce them have lead to a wide range of awareness-support tools that aid potential recipients to discover relevant material and originators to make material easier to discover.

### 5.5.1 The Time Dimension in Virtual Cooperative Interaction

Awareness and coordination of cooperative interaction involve the processes of *social entrainment*. This section examines the relationship between the time cycle of virtual cooperative interactions and the relative strength of extensional and intensional awareness.

When two or more individuals participate in virtual cooperative interaction, they often take on dual roles of originator and recipient in punctuated discourse. Gradually they become locked into *social entrainment* processes. Computer-mediated environments, such as newsgroups (Resnick et al, 1994) and shared drawing systems (Ishii & Kobayashi, 1992), provide specific external signals which set the pace of virtual cooperative interactions for participants. For example, the average *time cycle* for posting to a newsgroup and receiving a feedback ranging from one to a few days. Whereas the partial time cycle (t1+t2) for moving a mouse cursor in a real-time shared drawing system is around one to ten seconds. *External entrainment* occurs when the *actual* time cycle of a virtual interaction falls into the range of the *expected* time cycle anticipated by individual participants. When there is a wide discrepancy between the expected and the actual time cycle of interaction, participants often feel frustrated and decrease their desire to interact. For example, if cursor movements in a shared drawing system begin to take more than a few seconds to complete, the participants will tend to stop their interaction.

Continuation of virtual cooperative interaction can also break down when mutual and external entrainment processes are not synchronized with one another. When co-reviewing a book, the natural time cycle for mutual entrainment is in days and weeks, since it often takes that amount of time to read a book and absorb the material properly. It is unlikely that co-reviewers will want to use Internet Relay Chat (Reid, 1991) to disseminate and exchange their reviews. Such a fast time cycle of interaction is not well suited for activities involving deep, reflective cognitive processes.

The relationship between the time cycle and the relative strength of *extensional/intensional awareness* in virtual cooperative interactions can be illustrated in a time-dimension diagram (Figure 13). If the time cycle is relatively short, say a few seconds or minutes, we have an interaction that can be characterized as synchronous (real-time). If it is longer, we have an interaction that is often described as asynchronous (delay-time). The key notion here is that the types of virtual cooperative interactions are differentiated on a temporal continuum rather than by discrete categories.

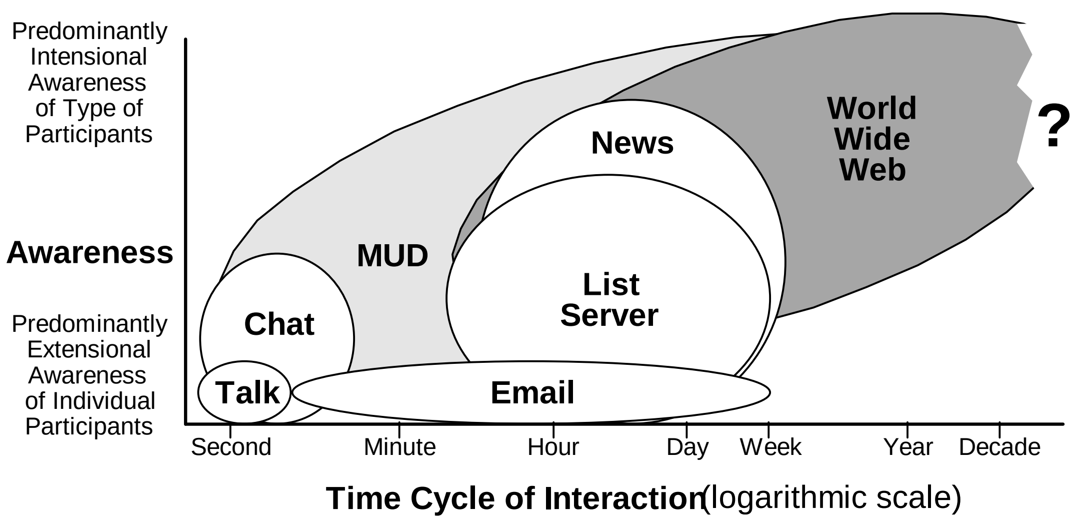


**Figure 13 Time Dimension of Virtual Cooperative Interaction**

In intensional oriented interactions, the level of intensional awareness is relatively high compared to extensional awareness; whereas in extensional oriented interactions, extensional awareness predominates. Many groupware systems (e.g., co-authoring systems, shared workspace systems) have been designed to support collaborative teams in which interactions are between known group members. Therefore, in these computer-mediated environments, cooperative interactions focus on extensional awareness. In contrast, interactions in USENET newsgroups involve both extensional and intensional awareness of a targeted audience. For example, one can respond to a question from a specific individual (an extensional oriented interaction) but do so publicly with the intention to address others who may have a similar question in mind (intensional oriented interaction).

The time cycle in virtual cooperative interaction often varies according to the cognitive processes involved in any given moment of an activity. For example, during a collaborative writing session (Neuwirth et al, 1994): when the co-authors' focus is on correcting sentences or paragraphs, the time cycle involved is usually around few minutes; and when they focus on reviewing chapters, the time cycle involved shifts to hours. Therefore, co-authoring systems are classified in the range of time cycles from seconds to days, in addition to be extensional oriented.

The time dimension diagram of virtual cooperative interactions allows us to visualize CMEs in terms of an *interaction area* they encompass as shown in Figure 13. The area denotes the range of a time cycle and the degree of intensional vs. extensional awareness.

The figure shows *talk* and *email* involving extensional awareness of the individuals involved but with the email cycle time being longer, corresponding to *talk* being 'synchronous' but *email* 'asynchronous.' The figure makes it clear that it is important to think of the degree of synchronity as an analog dimension not as a binary distinction. *Chat* is shown as overlapping talk but usually involving some lesser awareness of other participants and longer time cycles in discourse.

*Listservers* and *newsgroups* operate on time cycles of hours to days and involve extensional awareness of only a few participants but a reasonably strong intensional awareness of the type of participant. The web operates on even longer time scales from a day up to years and generally involves little extensional awareness of participants—readers may know particular authors, but writers generally do not know their readers. The question mark indicates that the web has been operating for such a short period that any long term estimates of its impact as a publication medium are speculative projections.

MUDs (Bruckman, 1994; Bruckman & Resnick, 1993), as noted previously, are anomalous in combining chat- and web-like services, and span a major part of the plot. This may seem strange in that they are not ubiquitous like the other services shown, although there are significant uses of MUDs to support professional communities (Evard, 1993). However, web browsers are being extended to integrate the chat facilities of MUDs as well as providing access to email, news, new collaborative tools and audio-visual interaction. Hence, as happened to Gopher, the web is subsuming MUDs and the entire spectrum of services shown in Figure 13 will appear to the end-user as an integrated service. However, the human factor distinctions between the different forms of discourse being supported will remain (Gaines, Chen & Shaw, 1997).

## *5.6 Awareness Structure Involving Virtual Cooperative Interaction*

We can regard the global cyberorganism as a virtual community of subsystems (themselves being living systems). These subsystems or agents provide resources to one another with the most significant dimension relating to overall system coordination being that of the *awareness* of who is providing a particular resource and who is using it.

Many workflow processes in team-work environments share a common goal of trying to assist people to work collaboratively as a cohesive team and to provide them with a sense of common purposes (e.g., completion of the group task). A critical requirement in shared tasks is maintaining *situational awareness* by keeping everyone adequately informed (Norman, 1993).

### 5.6.1 Collective Awareness Taxonomy

As we have seen previously in the last chapter, the *collective intelligence* model (Smith, 1994) focuses primarily on project groups and/or relatively small, close-knit teams within larger groups in designing/creating artifacts (e.g., committee reports, computer software). This comprehensive model takes into account cognitive processes within individuals, artifacts to-be-built, and situated actions.

Using the human self-awareness model and cognitive Information Processing System (IPS) architecture (Newell and Simon, 1972; Newell, 1982) as references, two analogous forms of awareness can be identified for collective groups: awareness of the group's collective long-term memory and awareness among and of each other.

The model identifies three forms of awareness with respect to a group's Long Term Memory:-

1. close, detailed, *deep awareness* of a particular segment of the artifact;
2. less detailed, but still substantial, *peripheral awareness* of the artifact's adjacent parts;
3. the much thinner, *global awareness* of the artifact as a whole that is shared by the entire group.

A different kind of awareness is the awareness members have of one another. This category of awareness is closer to the notion of situational awareness described earlier. The model identifies three forms of awareness among members of one another:-

1. cognitive-guide-map like, *resource awareness* for locating specific knowledge and expertise among the group members;
2. emotional, but rationalized, *task-socio* awareness concerning social and political dynamics within the group in relation to the artifact/task.
3. and more situational, *chronological awareness* about when, what, and by whom something in the collective system is changing.

The above collective awareness taxonomy provides a useful starting point for investigating awareness maintenance mechanisms involved in supporting virtual collaboration on the Internet.

### 5.6.2 Awareness Hierarchy in CyberOrganism

The cyberorganism framework focuses mainly on the three levels of coordination behaviours above the organism level in the living systems theory: group, organization, and community. They are presented as the shaded areas in Figure 3. It is used to sketch out an investigative framework for studying virtual cooperative interaction. The Internet community at large, the virtual special interest communities, the virtual collaborative teams, and the virtual persona of an individual/agent can be considered as awareness hierarchy within the global cyberorganism. The four levels of the living systems theory nicely mirrors the levels in the awareness hierarchy in the global cyberorganism (Table 2).

| **Living Systems Level** | **Awareness Hierarchy** |
|---|---|
| Community | The Internet Community at Large |
| Organization | Special Internet Community |
| Group | Team |
| Organism | Individual |

**Table 2 Living Systems Level vs. CyberOrganism Hierarchy**

In a tightly-coupled **team**, each individual is usually aware of who will provide a particular resource and often of when they will provide it. In logical terms, this can be termed *extensional awareness* because the specific resource and provider are known, as contrasted to *intensional awareness* in which only the characteristics of suitable resources or providers are known.

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 virtual community. Instead, resource providers usually have an *intensional awareness* of the resource users in terms of their characteristics as *types* of users within the virtual community. The classification of users into types usually corresponds to social norms within the virtual community, such as the ethical responsibilities in a professional community to communicate certain forms of information to appropriate members of the community. Resource users in a special interest virtual 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 asymmetry between providers and users characterizes a special interest virtual community and also leads to differentiation of the virtual community in terms of core members of whom many users are extensionally aware, and sub-communities specializing in particular forms of resource.

In **the Internet community at large**, there is little awareness of particular resources or providers and only a general awareness of the rich set of resources is available. Awareness of the characteristics of resources and providers is vague, corresponding to *weak intensional awareness*.

These distinctions are summarized in Table 3, and it is clear that the classification of awareness can lead to a richer taxonomy of communities than the 3-way division defined. Analysis of awareness in these terms allows the structure of a virtual community to be specified in operational terms, and in complex communities there will be complex structures of awareness. The coarse divisions into sub-teams and sub-special interest virtual communities provides a way of reducing this complexity in modeling virtual communities.

| | **Team** | **Special-Interest Community** | **Community at Large** |
|---|---|---|---|
| **Resource Provider** | Extensional awareness of actual users. | Intensional awareness of types of users. | No awareness of users, or only weak intensional awareness of types of users. |
| **Resource User** | Extensional awareness of actual resources and providers. | Extensional awareness of actual resources and providers, or intensional awareness of types of resources and providers. | No awareness of resources or providers, or only weak intensional awareness of types of resources and providers. |

**Table 3 Awareness Hierarchy**

The differentiation of communities in terms of awareness draws attention to the significance of supporting various aspects of awareness in a living system component.

*Resource awareness*, the awareness that specific resources or resources with specified characteristics exist, may be supported by various indexing and search procedures. However, there is also a need to support *chronological awareness*, the awareness that a resource has changed or come into existence (Chen & Gaines, 1997a). In the next chapter, Table 5 will show the way in which current mechanisms for awareness support are classified by the awareness hierarchy.

## *5.7 Motivations for Virtual Cooperative Interaction*

This section examines the motivational dimension of virtual cooperative interaction. Here, a theory of *collective social exchange* attempts to explain the behaviors of participants in terms of exchange theory, effects of norms in virtual cooperative interaction, capacity of power and social influence. These motivational explanations together with *social learning theory* (in the next section) examine the fairness and reinforcement issues involved in virtual cooperative interaction.

When many individuals participate in a multitude of punctuated discourses (Figure 12), a chain reaction occurs. The accumulative effect generated by this chain of inquiry-response-reaction-response-reaction (and so on) is an evolving topical thread that can become a part of *shared knowledge* (Smith, 1992) among members of a virtual community. Through automatic archival services such as *Hypermail* (EIT, 1994) or some individual efforts such as FAQs (Frequently Asked Questions) and web pages, the shared knowledge persists and grows. An interesting question is: why should individuals contribute to this pro-social process? Correspondingly, how does the virtual community ensure its participants to contribute to the growth of the knowledge pool?

First, why would individuals want to participate in virtual cooperative interaction? Generally, interpersonal behavior can be characterized as a *social exchange* between people, and these social exchanges typically involve both rewards and costs to participants. On a balance, an individual will perform those actions which produce the greatest rewards at the least cost (Shaver, 1987). Therefore, according to this cost-benefit calculus, a perceived potential for rewards must exist for individuals to participate and contribute in a cooperative relationship.

In contrast with *classical social exchange theories* (Cook, 1987) (e.g., Kelly and Thibaut's (1978) *Interactional Matrix* model) which emphasize dyadic interactions between individuals, the **collective social exchange theory** focuses on interactions between individuals and their virtual community. Conceptually, the Internet community is viewed from a *collective stance* (Gaines, 1994b) as an entity to 'whom' individual participants exchange information resources with. This collective entity offers participants a valuable informational service—namely, as *a pool of human knowledge* (Berners-Lee et al, 1994)—in exchange for their contributions.

The *norm of reciprocity* is fundamental to social exchange and leads to contributing behavior. The reciprocity norm creates an obligation for repayment that must be satisfied if the interaction is to continue (Shaver, 1987). However, the way reciprocity operates in collective social exchange is more subtle than in conventional social exchange between individuals. Why should one reciprocate (through contribution) in a situation where social responsibility is relatively *diffused* among members in a virtual community?

One motivation for contributing to the net is for an individual to gain a positive *self-image* (Jones & Pittman, 1982). In this case, an individual has internalized the norm of reciprocity and acts according to the principle of **equity theory**: that is, a person will seek to maintain his ratio of rewards to costs as the same as that of relevant comparison persons (Walster, Walster & Berscheid, 1978). A sense of *guilt* would occur if the individual perceives he has not contributed enough to the virtual community. Hence, he would want to reciprocate fairly.

Another more subtle motivation is that of contribution as an investment in **social power**, that is, the *capacity* of a person or group to affect the behavior of another person or group (Schopler, 1965). Contributions made by an individual may not only help others, but may also help her to gain name recognition from peers. The more one contributes publicly and receives recognition for one's contributions, the more one gains the capacity of power to influence others or the virtual community as a whole. The added weight in recognizing who is first to contribute relevant information also motivates individuals to volunteer information resources more readily. As seen earlier in Chapter 2, the competition for priority in contribution has been well documented in Merton's studies on the reward system in scientific discovery (Merton, 1973).

The motivational dimension illustrates the importance of initial conditions for *cybernetic feedback loops* (Miller, 1978) in reinforcement of virtual cooperative interaction. It provides a coherent explanation for the apparent altruistic behavior of information providers on the web/net.

## *5.8 Reinforcement of Virtual Cooperative Interaction*

One question raised earlier in the dissertation is that: "why do people publish information resources on the web in the first place?" Usually a resource provider might never know the identity of her resource user; nevertheless, she contributes even without any potential and apparent pay backs for her effort. Two possible motivations described earlier for providing information resources on the web are gaining positive self-image and name recognition. How does such a pro-social behavior initiate and continue?

The concern here is with the relationship between the *effect* of an individual's behavior in a *virtual community* and its impact on the individual's later behavior. This is the basis to **operant conditioning**, the learning process by which behavior is modified by the consequences of previous similar behavior (Ritzer, 1992). An individual emits some behavior. The virtual community in which the behavior occurs in turn "acts" back in various ways. The reaction—positive, negative, or neutral—affects the individual's later behavior.

**Social learning theory** suggests that novel social behavior is first learned through imitation of actions taken by others who act as (social) models (Bandura & Walters, 1963). The reinforcement received by a model serves as information to the person about which behaviors are acceptable and appropriate for the circumstances. Once a novel action has been acquired through imitation, its probability of continuation is depended on the reinforcement it receives. *Vicarious reinforcement*, as well as direct reward or punishment, can play a part in social learning (Shaver, 1987).

On the web, an individual's first successful encounter with a home page full of relevant information resource provides a positive role model for imitation. Her subsequent positive net-surfing experiences will further increase her exposure to other positive models. Once an individual internalizes the web culture which encourages construction of a personal home page (which coincidentally also provides a virtual persona for self-image), she will come to view that contribution to the web as pro-social behavior and act accordingly. The dynamic of *social exchange* then comes into play here, if the costs of putting up information resources (e.g., research papers, hyperlinks to relevant web pages) are relatively low to her (e.g., she has necessary skills and resources), she would contribute to the web. In addition, an original intention to contribute to the web community does not need to exist, she may be coincidentally using her home page to organize her knowledge resources and contribute to the web community as an after thought (or as a by-product). In this situation, the extensional audience is herself together with a vague sense of intensional awareness of other potential resource users.

The web/net culture itself is a product of the *positive feedback process* (Miller, 1978). The initial cultural memes have been seeded by the scientific ethos, value systems, and social norms as described in Chapter 2. As more people start to use the net and become parts of the global cyberorganism, they are initiated into the culture of virtual cooperative interactions. These culture memes are like *viruses of the mind* (Dawkins, 1989b; Dennett, 1995) invading the new inductees' conscious. As more people exhibit the virtual cooperative interactions (pro-social behaviours), the probability of new users encountering proper social models equally increases; hence, the higher probability for successful social learning. The cumulative results of successful social learning are fed back into the social system within the cyberorganism. Eventually this positive feedback phenomenon for inducting new members become self-perpetual.

How does reinforcement come into the picture? Frequently, one would encounter some home pages that had been constructed months or years ago without any revisions or new contributions. Their authors have neglected them and ceased to contribute. Once a novel behavior has been acquired, it needs to have *intermittent, positive reinforcements* to sustain the behavior (Bandura & Walters, 1963). In order for reinforcement to take place, there must exist a feedback loop. The round-trip cycle of virtual cooperative interaction provides an individual the necessary awareness of the *effectiveness of investment in social power* which is crucial to reinforcing the behavior and leading to similar future actions.

An observable measurement of the effectiveness of social power on the web is the relative popularity of a web site. The popularity of a home page can be inferred from recognition earned by its page access frequency counter, commentaries in its public guest-book, awards given by reviewers of popular web sites, and the number of other web pages linked to the page, etc. These gauges of popularity (which measure the relative power for social influence) provide direct reinforcements (can be either positive or negative) to an information provider. They also offer indirect, vicarious reinforcements to other information providers by providing social models for comparisons.

## *5.9 Summary*

This chapter presents a conceptual model for virtual cooperative interaction within the global cyberorganism. The model encompasses the communication processes and collaborative knowledge acquisition activities from closely-coupled teams to those of the very diffuse Internet community at large. It analyzes these activities in terms of the punctuated discourse processes, breaking down the cycles of action and response involved into a continuous temporal dimension. It analyzes them also in terms of awareness by originators of recipients and vice versa. The temporal dimension and awareness hierarchy enable the existing taxonomies and models of CSCW to be extended to encompass a very wide range of systems operating in both the short- and long-term and ranging from small teams to large communities. The model analyzes motivational aspects of virtual cooperative interaction. It gives rise to natural structural analyses of the activities which allows the types of communities involved to be identified from their observed activities. It can also be used to categorize computer-mediated environments roughly into groupware and socioware. Their respective general characteristics are listed in Table 4.

The five elements model presented in the chapter implies that for successful maintenance of continual virtual cooperative interaction, the following criteria must exist within the cyberorganism:

- establishment of resource awareness for initial encounter
- establishment of mutual awareness as a feedback loop for continual virtual cooperative interaction
- compatibility between the expected and the actual time cycles of virtual cooperative interaction
- properly situated expectation of fairness in terms of collective social exchange
- accumulation of favourable feedback for reinforcement in virtual cooperative interaction

| | Groupware | Socioware |
|---|---|---|
| **Primary Purpose** | team, task-oriented collaboration | ($2^{nd}$ sense) ‘virtual’ cooperative interaction |
| **Awareness** | strong mutual<br>extensional | weak mutual<br>intensional |
| **Time Cycle of Interaction** | short to medium<br>(seconds to days) | medium to long<br>(hours to years) |
| **Motivation for Cooperation** | individual social exchange | collective social exchange |
| **Power Relations** | well-defined roles as part of team definition | emergent roles from investment in social power capacity |

**Table 4 Comparisons of Groupware and Socioware**

The current model also identifies the types of groupware and socioware that are needed to expedite collaborative and virtual cooperative activities, and provides a framework for classifying existing tools in use on the Internet. It focuses on participants’ motivations and power relationship which determine their social roles, goals, expectations in virtual cooperative communities. Those social constructs are generally implicitly defined in groupware by the nature of group tasks (Mandviwalla & Olfman, 1994) and organizational structures (Kling, 1980), but are emerged, self-organized through socioware.

This chapter thus concludes Part I of the dissertation. From chapter 1 to chapter 5, I have presented the fundamentals of the cyberorganism framework: living systems concepts; origins and evolution of the net; structures and processes in its critical subsystems; control, coordination and awareness issues; and elements of virtual cooperative interaction.

The next chapter begins Part II of the cyberorganism framework. The subsequent chapters are concerned with systematic techniques and methodologies. Together they investigate the cyberorganism framework's utility, postulate investigative and methodological questions derived from the basic framework. They layout the demonstrations of the utility of the cyberorganism framework and correspond to a logical flow from teams, to special interest communities, to that of the Internet community at large.

# CHAPTER 6

# Awareness Maintenance Mechanisms

One key characteristic of living systems is that they maintain steady states through feedback controls. Each subsystem or component within the collective system needs to be aware of each other's state information critical to their cooperative tasks at hand. Here the timer subsystem is critical in providing coordination information in order to achieve both *mutual and external entrainment* processes. Awareness maintenance mechanisms provide the essential state signals in supporting *social entrainment* among subsystems or components.

This chapter examines human factor issues relating to awareness of changes on the web. The main focus here is on chronological awareness as a special case of situational awareness, that is, the awareness of when something (an event or an artifact) has changed (Chen & Gaines, 1996b). Some techniques for supporting chronological awareness in the cyberorganism manifested themselves *as awareness maintenance mechanisms*. The chapter will begin with a methodological framework for classification and evaluation of awareness maintenance mechanisms. Afterward, I will describe CHRONO and related awareness maintenance mechanisms in detail, then evaluate them using the framework. Finally, the framework will classify them and other awareness techniques on the Internet.

## *6.1 Dimensions of Awareness Maintenance Mechanisms*

Web pages, FTP archives, listservers, and other common Internet infrastructures are the primary means for information dissemination for collaborators within the cyberorganism. Sometimes these infrastructures are being constantly updated to reflect individuals' current states of knowledge on their portions of collaborating tasks. In a dynamic environment where large amounts of information are created and updated frequently, the need to keep up with the most up-to-date and relevant information has become more important as the Internet community expands.

### 6.1.1 Dimensions of Awareness Maintenance Mechanisms

There are four main dimensions of design considerations for awareness maintenance components:

| | | |
|---|---|---|
| 1. | Locus of Responsibility: | *Server-Side*, *Client-Side*, or *Centralized Dispatcher*. |
| 2. | System Hierarchy: | *Group*, *Special Interest Community*, or *the Internet at Large*. |
| 3. | Method of Locating Changes: | *Browsing* vs. *Targeting*. |
| 4. | Complexity of User Interaction: | *Simplicity* vs. *Customization*. |

The first dimension, the locus of responsibility, differentiates who is responsible for maintaining the record-keeping mechanisms for supporting awareness maintenance.

- A *server-side* approach ensures that only the users who are currently visiting the web site will need to know what information has been changed. Hence, it reduces network traffic by avoiding needless broadcasting of chronological information to some users who might not be concerned with it. The main disadvantage of this strategy is that in order to know whether or not any particular page has been changed, a user will need to check out the specific web site periodically.
- A *client-side* approach periodically monitors specific pages at various web sites and report whether or not they have been changed recently to ensure the user will be aware of any changes. The main disadvantage is that the users need to remember to run such a mechanism, or it must be set up to run periodically, in turn causing it to consume higher network bandwidth.
- A *centralized-dispatcher* approach put the monitoring responsibility at some specific central registry locations that automatically monitor the registered pages for the users.

Its main disadvantage is the high network traffic involved in such a centralized broadcasting scheme.

The first dimension can also be considered from an information resource provider/user perspective. Various information systems, such as the web, use the client-server model to partition the computational division of labor. Similarly, the locus of responsibility of awareness maintenance at every level can be divided into originators (i.e., providers) of information, recipients of information resources, and intermediaries of information retrieval exchange. Therefore:

- An *originator* is a source/server of information dissemination.
- A *recipient* is an user/client of information resource.
- An *intermediary* is a meta-information resource mediating between originators and recipients.

The second dimension, the awareness hierarchy, reflects the need for maintaining mutual awareness among members existing in various collaborative arrangements. There are three main levels of awareness arrangements which constitute the awareness maintenance hierarchy.

- The *team/group* level usually consists of closely coordinated members in a relatively small or medium size project. There is usually a need to be aware of short-term changes in the data being managed by different group members as part of their task activities, and the nature of this need is relatively well-defined.
- The *special interest community* level frequently involves a loose coordination of different group projects within an organizational structure such as a scholarly sub-discipline. There is usually a need to be aware of significant changes in the knowledge available in the community, and satisfying this need involves mechanisms ranging from well-defined channels, such as electronic journals or reprint archives, to organizational/cultural/ethical norms on how information is to be disseminated.

- The *Internet at large* level often involves mechanisms for providing its members at large some form of resource awareness, for example in locating where a specific information resource is on the web.

Viewed from the cyberorganism perspective, situational awareness is essential at every level above the individual level in the systems hierarchy and that originators and recipients of information are situated at the opposite ends of the channel and net information subsystems. The various awareness maintenance mechanisms (such as the chronological awareness support mechanisms for groups) serve as (or provide the functionality of) the timer, associator, and memory subsystems in various system levels.

The third dimension, the method of locating changes, involves two different ways of locating documents that have been changed: browsing and targeting.

- A *Browsing* approach facilitates the chronological browsing characteristic: visitors of a site may find relevant information via browsing the concurrently created/modified web pages, because closely related documents are sometimes created (or modified) around the same time. This browsing approach allows the “at a glance” attribute for accidental discovery of relevant information without prior awareness of their existence.
- A *Targeting* approach focuses on specific pages or information that users have previously specified. Therefore, this method of locating changes is more direct and efficient. However, the main disadvantage of such an approach is that the users cannot be made aware of any new information which have been created recently. They are limited only to changes made to prior knowledge.

Finally, in the fourth dimension, the complexity of user interaction, denotes the mechanism’s usability in terms of simplicity vs. customization.

- A *Simplicity* approach tries to make its user interface simple and familiar to web users. The goal is gearing toward ease of use and a shallow learning curve. On the web this is typified by mechanisms that present awareness information through the generation of familiar web documents.

- A *Customization* approach, in contrast, tries to allow elaborate customization of features, but it also demands more efforts by its users to learn and utilize its functionality. On the web this is typified by mechanisms that use separate tools with their own user interfaces to support awareness.

### 6.2 Chronological Awareness Support Mechanisms

This section presents a concrete example of a class of awareness support mechanisms specially designed for supporting chronological awareness maintenance on the web. One such mechanism is described in detail, followed by brief overviews of three other mechanisms.

By definition, a chronological awareness support mechanism for the web provides each individual an appropriate awareness of relevant activities of other individuals. Hence it allows team members to synchronize their activities in a more coherent way by keeping them informed and aware of any changes made to each other's web pages or other information resource that might be relevant to their current tasks. Maintenance mechanisms supporting such awareness correspond closely to the timer critical subsystems in a cyberorganism.

CHRONO is an HTTP server-side mechanism I implemented for the Knowledge Science Institute web server. It generates chronological listings of web pages recently having been changed or newly created. It provides a basic awareness-support that lets visitors of a web site (e.g., members of a group, an organization, or other net surfers) see which web pages have been modified since their last visit (Chen, 1996). Currently the CHRONO mechanism has been implemented for the UNIX platform and made widely available for use at other sites.[3] As shown in Figure 14, CHRONO presents to the visitors an HTML

[3] For CHRONO example, see http://www.cpsc.ucalgary.ca/~lchen/cpsc.html#chrono

document that lists the titles of web pages at the site in reverse chronological order. This chronological listing of web pages also functions as a collection of hyperlinks to the listed pages.

### 6.2.1 CHRONO System Implementation

The current mechanism has been successfully compiled for the SUN OS 4.1.3 and Silicon Graphics (SGI IRIX 5.3) UNIX platforms, but it should be easily portable to other operating systems. CHRONO is written in the C programming language and uses the standard UNIX file I/O functions (found in the standard C library). It traverses through the www/ and its sub-directories of a targeted website to generate a reverse chronological list.

The mechanism consists of a background process that is invoked by the UNIX system daemon *cron* on an hourly base. When invoked, this process recursively enters any pre-specified www/ (or public_html/) directory that comprises a web site. It looks for files with filenames (e.g., .HTML or .GIF) matched in an optional *global pattern-matching* file. By default, it seeks out HTML files, then extracts their title field and records their time/date of last modification. If a document does not have a title field, its file name is used instead. The process sorts the collected titles into reverse chronological order. Finally, it produces an HTML document containing the listing. The result is a master web document containing the hyperlinks to these periodically generated chronological entries. A new listing page is generated via the background process for the targeted web site every hour.

Embedded in each `<A HREF="url">` tag entry is a special time-stamp specified in the field: `LAST_MODIFIED="seconds"` (where the *seconds* value is the standard UNIX *utime*)  This field value is then utilized by META-CHRONO (a meta-level awareness support mechanism to be discussed later) to sort entries from multiple CHRONO listings.

Website maintainers have the option of  placing a *local-control* file to tell CHRONO which file/directory patterns to exclude (or to include) in indexing of specified sub-directories. This feature is useful if there are some sensitive documents or sub-directories, the maintainers wish to keep private from others. It basically prevents the creation of automated hyperlinks by CHRONO for those sensitive items.


**Figure 14 CHRONO In Use At a PC User Group Website**

### 6.2.2 CHRONO User Interface and its Functionality

The user interface of the CHRONO mechanism is straightforward and intuitive for web users. It looks like an automated *what's new* page to the users. From the list, the visitors are able to tell at a glance what documents have been modified or created recently. They can also scroll down the list to check those older documents at the site. Because the titles of the listed pages also act as hyperlinks to the actual web pages, visitors can simply click and jump to the relevant pages of interests.

Chronological hyperlinks presented in the CHRONO listings provide the visitors the means to access the newly modified or created pages. This time-line (or history) dimension complements the functionality of the associative memory characteristic found in typical hyperlinks which join related information.

The time-line dimension allows frequent visitors of a web site an immediate awareness on what have been changed since their latest visit. The changes may reflect some web pages in which they have been previously interested or they may show some pages that the visitors have never seen before but now appeal to them. Hence, this chronological browsing characteristic is analogous to spatial (subject-category) browsing-characteristic to what library patrons have often experienced when looking for books on open book-shelves (i.e., accidentally finding other relevant books near the books that they were looking for originally).

What is different here is that instead of finding relevant information via browsing the near by subject-categories, now the users may find relevant information via browsing the concurrently modified/created web pages. Sometimes, conceptually related documents are created (or modified) around the same time; however, their author(s) may not remember to update the HTML links to them. Unlike a manually updated *what's new* page in which the users have to rely on the timely updates made by a Web-master (or by the document authors), CHRONO provides the time-line dimension to the users automatically, in a reliable, periodic fashion.

### 6.2.3 Usage of CHRONO

The CHRONO mechanism has been running at the Knowledge Science Institute in Calgary since March, 1995. It originally serviced seven web locations: two research units and five individuals. People associated with the two research units: Knowledge Science Institute and GroupLab periodically had utilized the mechanism to check on new developments of each other (both within group and between groups).

From a preliminary examination of the HTTP access_log of the website which CHRONO monitored and from talking with individual group members, I discovered that the chronological awareness supports of the five personal sites did in fact offer group members more focused chronological awareness about each other's online working patterns. Occasionally, some people would discover new projects others were working on that they were not previously aware. The preliminary positive usage experience of CHRONO suggested to me that the mechanism would be of interest to outside research groups and websites.

Currently the CHRONO has no means for recording when a particular visitor last visited the site, so it relies mainly on the users' own recollection. Future versions of the mechanism may incorporate the ability to provide customized information to the users.

Eventually, CHRONO had been in use in over 60 different websites since its initial release in May, 1996. The diffusion processes of CHRONO and methodology for tracking them on the net are further examined in a later chapter.

Since CHRONO is now being used at a number of sites, a new mechanism—META-CHRONO—has been under development which collects and collates information from multiple sites running CHRONO and provide awareness of activities being carried out on a distributed basis (Chen, 1996).

META-CHRONO is a recent addition to support special interest community awareness for distributed research teams and groups (Gaines, Shaw & Chen, 1996). This meta-level awareness tool collects CHRONO listings from related websites, then it generates a collective chronological awareness listing reflecting the changes at those websites. This allows visitors to become aware recent modifications of changes at all registered websites via a centralized META-CHRONO index-listing. This meta-level awareness support of CHRONO listings is a natural extension from the cyberorganism framework (Chen & Gaines, 1997).

## *6.3 Other Related Chronological Awareness Mechanisms*

*WebWatch* (Specter, 1995), *Katipo* (Newberry, 1995), and *URL-Minder* (NetMind, 1995) are other chronological awareness tools that track changes in specified documents. Those mechanisms are based on different design and implementation criteria. These are briefly examined in this subsection, followed by a comparative evaluation of current chronological awareness mechanisms.

### 6.3.1 WebWatch

WebWatch is a client-side chronological awareness mechanism for keeping track of changes in selected web documents. Given an HTML document referencing URLs on the web, it produces a filtered list, containing only those URLs that have been modified since a given time. The criteria used for filtering can be given as a global setting that applies to all URLs, or can be derived automatically, using the time of the user's last visit to the document, as recorded by the web browser in the user's local HTML (e.g., bookmark) file.

In contrast with the simple time-line listing strategy used in CHRONO, WebWatch stores its arguments in a parameter file. Once the users have customized the program to their needs, using its graphical front-end, they can have it run periodically in unattended mode.

### 6.3.2 Katipo

Katipo is another client-side chronological awareness mechanism built for Macintosh that shares many similar concepts with WebWatch. It reads through the Global History file maintained by some web browsers checking for documents that have changed since the last time a user viewed them. The basic difference between it and WebWatch is that it uses the Global History file as its reference for checking URLs, whereas WebWatch uses the Bookmark file.

### 6.3.3 URL-Minder

URL-Minder is a centralized mechanism that keeps track of resources on the net and sends registered users e-mail whenever their personally registered resources change. Users can have the URL-Minder to keep track of any web resource accessible via HTTP. It can be anything—not just web pages users personally maintain.

The URL-Minder mechanism keeps track of one web page, image file, or other Internet resource at a time. It tracks the actual HTML markup, binary contents, or ASCII contents of the URL they have submitted. If an HTML page includes a GIF or JPEG graphic, the URL-Minder then informs subscribers via e-mails when the reference to the graphic changes. It checks on users' registered URLs at least once per week, and informs users if it fails to retrieve their registered URL after trying twice.

## *6.4 Evaluations of the Awareness Support Mechanisms*

This section presents a comparative evaluation of the four chronological awareness support mechanisms discussed previously: CHRONO, WebWatch, Katipo, and URL-Minder. Each mechanism has its unique approaches for achieving chronological awareness support for web users and complement each other along four main dimensions.

The first dimension, the locus of responsibility, differentiates who is responsible for maintaining the record-keeping mechanisms for chronological awareness. For example, CHRONO is a server-side mechanism in which chronological listings are being updated and kept at the web server-side. Thus, CHRONO can be thought of as offering "*chronological awareness on demand*". WebWatch and Katipo, however, put the responsibility of maintaining chronological awareness on the client-side. Both client side mechanisms periodically monitor specific pages at various web sites and report whether or not they have been changed recently. Finally, URL-Minder requires its users to register at a centralized site.

The second dimension, the awareness hierarchy, signifies that all of the current chronological awareness support components are mainly focused at the team level (and to some extend at the special interest community level). Collectively, they are supporting the chronological awareness of information resources typified by a closely-coupled collaboration at the group level.

The third dimension, the method of locating changes, involves two different ways of locating documents that have been changed: browsing and targeting. CHRONO uses the browsing approach in order to facilitate the chronological browsing characteristic. Conversely, WebWatch, Katipo, and URL-Minder employ a targeting approach in which they are targeted on specific pages or information that users have previously specified.

Finally, the fourth dimension, the complexity of user interaction, denotes a mechanism's usability in terms of simplicity vs. customization. CHRONO and URL-Minder are in the simplicity category; their user interfaces are simple and familiar to web users (i.e., scrolling list of hyperlinks and fill-in form of URLs and e-mail address). They are geared toward ease of use and a shallow learning curve. Both mechanisms, however, have no capability for individual customization. In contrast WebWatch and Katipo allow elaborate customization of features, but they also demand more efforts by the users to learn and utilize their functionality.

Therefore, each of the chronological awareness support mechanisms examined so far have various degrees of advantages and disadvantages along the four dimensions. CHRONO has the advantages of: (i) simplicity of user interface; (ii) supporting accidental discovery via its browsing characteristic; and (iii) server-side chronological awareness information on demand. It is nicely complimented by WebWatch and Katipo for their strength in the efficiency of targeting approach and customization capabilities. And finally URL-Minder offers another unique service: it uses e-mail as its notification channel. This approach is useful for users who use their e-mail services more frequently than web browsers. Together as a whole, these chronological awareness support mechanisms have covered a wide range of approaches in respect to four major dimensions of chronological awareness support.

## *6.5 Classification of Awareness Mechanisms*

The methodological framework developed in this chapter can be used to study a wide range of general awareness mechanisms for the web. There are various awareness maintenance artifacts on the net that address different system levels. The following analysis examines and categorizes them along the two major dimensions: *level of awareness hierarchy and locus of responsibility*. Currently, *intermediary* mechanisms are still in a nascent stage of evolution (e.g., Universal Resource Agents), hence the present classification does not yet reflect the intermediary locus of responsibility.

| Locus of responsibility | Team/Group | Special-Interest Virtual Community | The Internet Community at Large |
|---|---|---|---|
| **Originator** | *Extensional awareness of actual recipients.*<br><br>Use *email* to notify.<br><br>Use **CHRONO** to index. | *Intensional awareness of types of recipient.*<br><br>Broadcast to *listserver*.<br><br>Establish *HTML links*.<br><br>Use **META-CHRONO** to index. | *No awareness of recipients, or only weak intensional awareness of types of recipients.*<br><br>Broadcast to *newsgroups*.<br><br>Register in **Yahoo**.<br><br>Initialize **Alta Vista**. |
| **Recipient** | *Extensional awareness of actual resources and originators.*<br><br>Use email to inquire.<br><br>Check **CHRONO** index.<br><br>Use **WebWatch**, **Katipo** or **URL-Minder**. | *Extensional awareness of actual resources and originators, or intensional awareness of types of resources and originators*<br><br>Subscribe to listserver.<br><br>Follow *HTML links*.<br><br>Check **META-CHRONO** index.<br><br>Use **WebWatch**, **Katipo** or **URL-Minder**. | *No awareness of resources or originators, or only weak intensional awareness of types of resources and originators.*<br><br>Read *newsgroups*.<br><br>Browse **Yahoo**.<br><br>Search with **Alta Vista**.<br><br>Search with **MetaCrawler**. |

**Table 5 Classification of Awareness Mechanisms**

At the *team* level, originators of the information resource can organize and implement work flow models of the group activities, use server-side chronological awareness support mechanisms such as CHRONO, or/and send e-mail notification to users of information. Recipients of the group level information resource can use client-side chronological awareness tools such as WebWatch and Katipo or register in a centralized dispatcher service like URL-Minder. Alternatively, they can send e-mail to inquire to information originators to see if any new things have come up.

At the special interest community level, originators of the information resource can broadcast to concerned individuals, groups, or organizations via specific listservers, or announce in organization-maintained MOO or MUD. They can also establish what’s new HTML links in organization news while recipients can participate in HyperNews or MOO, and follow the new HTML links in organizational web pages.

At the net community at large, originators can register the information resource in hierarchical subject services like Yahoo or in NCSA's, What's New service, or initialize their pages in searching and navigational services like Alta Vista, while recipients can browse Yahoo, search Alta Vista, LYCOS, or read and post to the USENET newsgroups.

One interesting observation can be made about the level of awareness in relation to the level of coordination. As the level of awareness moves from the group level to the community level, the need for closely-coupled coordination decreases among members. What happens in practice is the awareness maintenance becomes asymmetrical, rather than mutual, at the higher system level. However, the major awareness requirement continues. Resource providers may not need to be aware of who their users are, but the users' activities may be critically dependent on the status of the resources.

## *6.6 Summary*

The web has grown very rapidly to become a major resource supporting collaborative activities in a wide range of groups, disciplines and communities in the global cyberorganism. However, the growth of the web creates problems of information overload and of maintaining awareness of activities at other sites relevant to one's own tasks. This chapter develops methodological dimensions for studying and supporting awareness on the web. It describes CHRONO, an awareness maintenance mechanism for providing a feedback channel for changes at subsystems or components. The last section uses two key dimensions of the methodological framework to classify CHRONO and related mechanisms, and to clarify the human factor design issues involved.

# CHAPTER 7

# Systematic Methodologies for Listserver Analysis

In chapter 4, we have seen that special interest communities—i.e., associator subsystems—are situated at the pinnacle of an integrated architecture for communication, social, and knowledge processes (Figure 8). As observed earlier in the five elements model of virtual cooperative interaction (Chapter 5), a mailing list server handles the primary discourse structure of a special interest community. In this chapter I will propose systematic methodologies for analyzing listservers on the net. One of the key objectives is to gather various types of useful statistics for gauging the life cycle and vitality of virtual communities on the net through observable interactions at the special interest community level. This is the level characterized by the *organization* level in the *living systems theory* (Miller, 1978).

## *7.1 Introduction*

Before I start describing the proposed methodologies for analysis listservers in detail, here is a disclaimer: the *only point of interest* here is what *happens overtly and observably on the net*. Although at first glance, it may seem that such a limited world view is rather inadequate. Some people may find the exclusion of many aspects in real life of an individual quite objectionable. Indeed there are so many things in a person's real life that can influence her thinking and behaviours, which must be at least equal, if not more important than online, overt actions. Personally I agree with such sentiments, nevertheless the aim of the current research is on modeling the net. Hence the subject for the current analysis methodologies is cyberorganism, not real life individual psychology or sociology. As stated by Popper (1972): the objective existence of *World 3* entities and spaces meant that we could examine them, evaluate, criticize, extend, and explore them, *in public*. The key words here are *World 3* and *in public*. Even though many off-line behaviours of people may be more important or may influence their online persona indirectly, the focus here are *overt behaviours* at the level of special interest communities in the cyberorganism.

The community structure of discourse through listservers may be studied through statistical analysis of the server archives, and shows that a small group of members generally dominate the discourse. Table 6 shows the number of items, number of different contributors, number of contributors with 5 or more mailings, and the group of authors who account for 50% of the total mailings to the list for the conceptual graphs listserver (Gaines, 1993b). This list commenced in August 1989 serves a community of some 220 researchers concerned with knowledge representation based on Sowa's (1984) work, and Sowa was himself a major contributor during its formative years. It can be seen that between 4 and 13 members of the community account for over 50% of the items mailed, and that this group is consistent over the 3 year period studied (Gaines, Chen & Shaw, 1997).

**Table 6 Patterns of Contribution to a Listserver**

| Year | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 |
|---|---|---|---|---|---|---|
| **Total number of items** | 270 | 325 | 607 | 720 | 1040 | 396 |
| **Total number of authors** | 37 | 63 | 125 | 168 | 213 | 108 |
| **Number authors with 5 or more items** | 14 | 13 | 24 | 27 | 49 | 22 |
| **Number of authors accounting for 50% of items and number of items for each of these authors** | 4<br><br>69 Sowa<br>34 Ellis<br>26 Tjan<br>17 Wei | 4<br><br>72 Sowa<br>58 Ellis<br>26 Tjan<br>17 Wei | 8<br><br>101 Sowa<br>83 Ellis<br>31 Esch<br>24 Tjan<br>17 Brkich<br>16 Bejan<br>16 Lehmann<br>14 Wei | 11<br><br>145 Sowa<br>47 Ellis<br>23 Hayes<br>22 Esch<br>21 Lehmann<br>19 Brkich<br>17 Levinson<br>14 Tjan<br>4 Willems<br>11 Mineau<br>9 Moulin | 13<br><br>126 Ellis<br>113 Lehmann<br>74 Sowa<br>55 Hayes<br>33 Dwiggins<br>22 Anquetil<br>21 Lukose<br>16 Brkich<br>16 Willems<br>15 Tjan<br>14 Wermelinger<br>14 Esch | 11<br><br>54 Sowa<br>47 Ellis<br>26 Hayes<br>13 Plotnikov<br>11 Esch<br>10 Delaguch<br>10 Lehmann<br>8 Moeller<br>8 Wei<br>8 Wermelinger |

This pattern of behavior shown in Table 6 is typical of most special interest communities where a relatively small number of facilitators both introduce topics and respond to most discussion items introduced by others. The small group has extensional awareness within itself and often acts as a team, whereas it has only intensional awareness of the types of background of other members on the list. There are strategies which allow facilitators to ensure that those posting discussion items have the satisfaction of seeing responses, yet attempt to reduce the dominance of the facilitators, for example by deliberately introducing a delay of a few days in their own responses to encourage others to respond instead of them. However, in some cases it is apparent that the person introducing a discussion item is hoping for a response from an extensionally defined particular person and this strategy would then be inappropriate.

The human factors of listserver interactions merit much more detailed study than they have received to date. Bales and Cohen’s (1979) SYMLOG methodology, based on Bales (1950) deep analysis of small group dynamics, provides an appropriate methodology. He classifies communications along three major dimensions: dominant—submissive, friendly—unfriendly, and instrumental—emotional. To these may be added a classification of the focus of discourse, whether to a specific individual or the group, and a content analysis of the memetic processes of developing a particular knowledge product.

Losada, Sanchez and Noble (1990) have used an extended SYMLOG methodology to classify detailed observations of small groups working together around a table, and have developed computational time series analyses which, presented in graphical form, clearly model the group’s behavior and show dysfunctional modes of operation. A similar approach to the analysis of listserver archives would provide empirical data on the human factors of discourse in a virtual community (Gaines, Chen & Shaw, 1997).

### 7.1.1 Methodological Assumption and Approach

A *virtual community*, for all practical purposes, is defined here by the membership of a special interest community (where people who share a common interest would subscribe intentionally). The membership criterion effectively creates a natural boundary for the delineation of a special interest community. The current research takes a cyberorganism stance in the sense that examining the interaction patterns and social network structures of a special interest community can tell us more about its nature and life cycle than simply analyzing the composition categories of individual members. A collective system is greater than the sum of its parts. The key to understanding this concept is to appreciate the difference between group-dynamics and individual psychology. The group-dynamics approach, by its name, implies that it focuses on the “dynamic” aspects of the group: interaction patterns and evolving network structures; this approach supplements and supplants individualistic methods in psychology and sociology. Typically those methods focus on the "categorical" and "static" aspects: such as members' individual attributes in relation to their group.

The proposed systematic methodologies share the ethos of social structural analysis: "*structured social relationships are a more powerful source of sociological explanation than personal attributes of system members*" (Wellman & Berkowitz, 1988). Many mainstream sociological studies treat social structure and process as the sum of individual actors' personal attributes. These attributes, whether derived genetically (e.g., gender, age) or socially (e.g., socioeconomic status, political attitudes) are treated as entities that the individual processes *as individuals*. Each is treated as an independent unit of analysis and lumped into social categories with others possessing similar attribute profiles. The method of analysis—such as: cross-tabulation, correlation, or more complex multivariate techniques—processes by sorting individuals processing similar combinations of attributes into similar analytic cells: for example, older men of high socioeconomic status who vote for Tory (Wellman, 1988).

Although statistical methods in sociology have grown increasingly sophisticated, they continue to focus on treating individuals as independent units. The very assumption of being statistically independent, which makes these methods so appropriate and powerful in categorical analysis, detaches individuals from social structures and forces analysts to treat them as parts of a disconnected mass. Such methods can only measure social structure indirectly, by organizing and summarizing numerous individual co-variations. They are forced to neglect social properties that are more than the sum of individual acts (Wellman, 1988).

Therefore, the current proposed methodologies take an integrative approach for analyzing virtual communities categorized by *communication patterns* on listservers. In addition to descriptive statistics, it utilizes social network and group-dynamics analyses. This approach can offer researchers more comprehensive and dynamic views of the growth, evolution of the studied virtual communities.

### 7.1.2 Integrative Methodologies

In terms of analyzing listserver interactions, two types of analysis exist: first is the content-free analysis of messages traffic in which the focus of the analysis is on the discourse patterns, not the actual contents of message; second is the analysis of group-dynamics in which encoded group behaviors can be captured through analyzing contents of messages posted by each list participant.

In both cases, the basic assumption of the current methodological approach is that most people on a particular list share interests on topics relevant to the list. Common interests would allow the self-selection process of the subjects into a particular special interest community to take place. People may have diverse backgrounds in terms of age, gender, personality, education, and so on. However individual differences on those diverse attributes are not considered here. The present methodology concentrates on the "observable" (overt) behaviours of individuals. In this case, one observable action is posting a message to the list (other actions: e.g., subscribing and unsubscribing to the list).

To determine the social network structure of the community, one can measure many *sequential statistics* (Bakeman & Gottman, 1986) from people's interactions on the list, such as: who posts to the list, who replies to whom, or which topic thread is growing. From which, sociomatrices, sociograms or directed-graphs (Wasserman & Faust, 1994) of the list can be determined and further analyzed. Finally, SYMLOG field diagrams (a system for the multiple level observation of groups, by Bales & Cohen, 1979) can be plotted after sequential event-coding of messages posted to the list. Together sociograms and SYMLOG field diagrams allow us to visual the group-dynamics more readily.

Various questions about the virtual community can be addressed from the current analytical methodologies, like: what is a pattern of the growth and vitality of the list community? How does its social network evolve? Who are the main contributors in the community? Who have the main influences in the community? How do such patterns emerge? Do they affect the vitality of the community? What are the psychological fields in the community and its sub-groups? Do polarization or unification result from social structures? Finally does polarization tend to create leaders and subgroups?

In order to answer those questions about a particular virtual community, we first need to gather measurements about the listserver.

### 7.1.3 Graphical Measurements

There are three types of graphical measurements available for a listserver. After the current analysis methodologies have been applied to a listserver, a global perspective about its communication dynamics and network structures can be inferred from those graphical measurements.

- Time-series plots of the message traffic flow.
- Sociograms and sociomatrices from the discourse patterns.
- SYMLOG field diagrams of the individual, clique, and community.

Below are brief descriptions about those measurements, more in depth coverage can be found in the following sections.

First a time-series plot of message traffic provides a chronological portrayal of the vitality of a community (Figure 15). One of the most basic measurements is the cumulative message count per specific interval (e.g., few days, a week) for a listserver. The time-series plot allows us, at a glance, to visualize the vitality of a special interest community. We can easily identify evolutionary patterns of the list from its time-series plot.

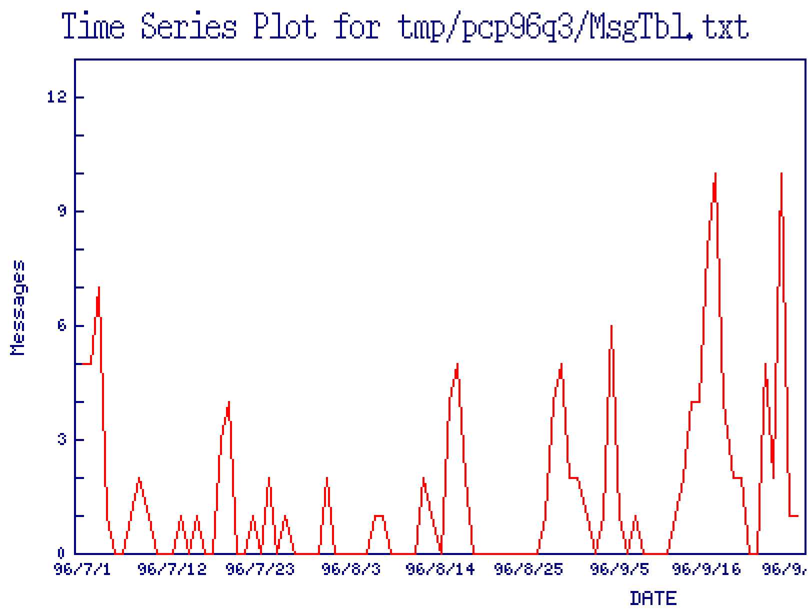


**Figure 15 Time-series Plot of Message Traffic**

A sociogram is a directed graph about the relationship-system between members on a list. This directed-graph can be derived from a sociomatrix of the discourse patterns on the list. For example, a hypothetical listserver with only 5 active members (among 10 subscribers). Their interaction patterns: who reply to whom publicly in a specific interval as be described in the hypothetical sociomatrix (Table 7). Each cell records the number of messages sent by member $M_{row}$ (the From row) to member $M_{column}$ (the To column).

| | To $M_1$ | To $M_2$ | To $M_3$ | To $M_4$ | To $M_5$ |
|---|---|---|---|---|---|
| **From $M_1$** | - | 4 | 3 | 0 | 0 |
| **From $M_2$** | 3 | - | 3 | 0 | 0 |
| **From $M_3$** | 0 | 5 | - | 5 | 0 |
| **From $M_4$** | 0 | 0 | 4 | - | 0 |
| **From $M_5$** | 0 | 0 | 1 | 0 | - |

**Table 7 Sociomatrix of Messaging Patterns**

From the above sociomatrix example, we then construct a corresponding sociogram. It offers a more intuitive, visual presentation of the list's network structure. By looking at Figure 16, we can easily determine that not every members have the similar degree of influence and contribution. In this case, member $M_3$ has greater influence than others, since everyone else is interested in her opinions and eager to reply to her messages.

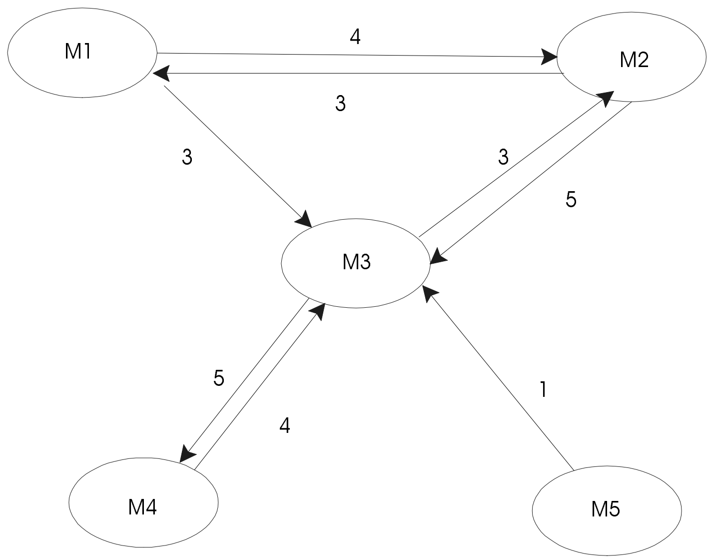


**Figure 16 Sociogram of the Network Structure of Table 7**

SYMLOG is a behavioural coding scheme operating within a three-dimensional space (Bales & Cohen, 1979). The three spatial dimensions map into three corresponding psychological dimensions: **Up-Down** (U-D dimension) whose psychological correlate is represented by dominant vs. submissive behaviour; **Positive-Negative** (P-N dimension), psychologically mapped into friendly vs. unfriendly; and **Forward-Backward** (F-B dimension), whose psychological correlate is task-oriented vs. emotionally expressive behaviour (Losada & Markovitch, 1990).

The main output of a SYMLOG encoded observation session is a "field diagram" which summarizes the average group behaviour by representing each participant as a circle whose radius conveys the level of dominance. The circle is located in a two dimensional plane whose vertical axis is the F-B dimension and whose horizontal axis is P-N dimension (Losada & Markovitch, 1990). By examining the field diagram in Figure 17[4],

---

[4] It is unrelated to sociogram in Figure 16.

we can see that Jane is the most dominant personality within the group. She also has the most friendly manner among the 5 members.

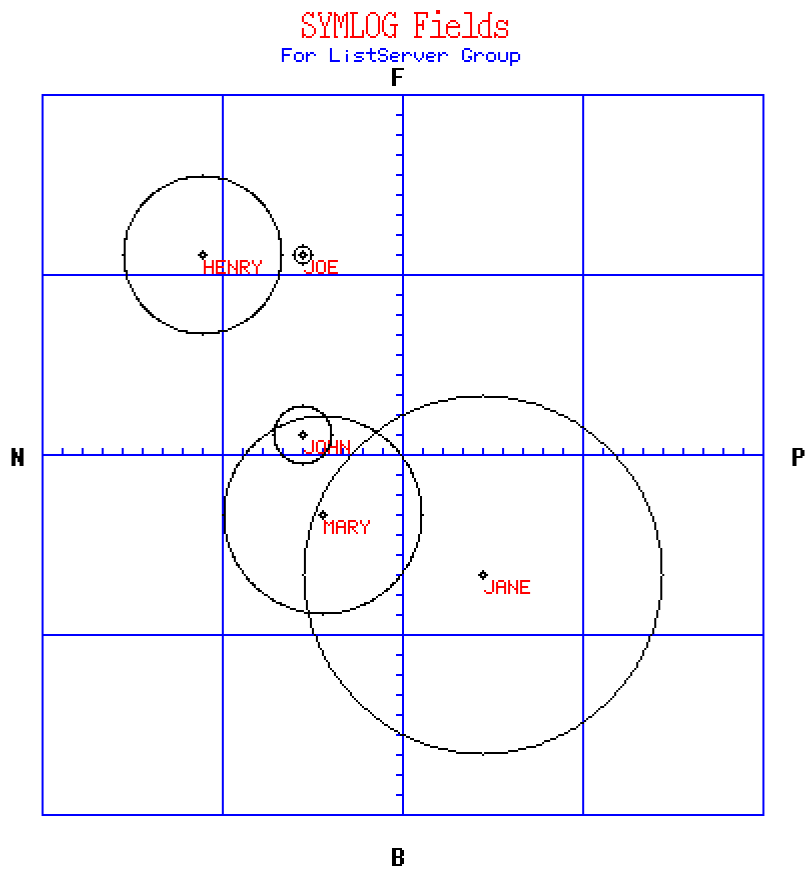


**Figure 17 Field Diagram of 5 Members**

Before collecting the above measurements, a systematic coding scheme is needed. Details about this coding scheme and the preparation process for encoding is presented in the next section.

## *7.2 Overall Coding Scheme*

The first step in observing interaction or archival analysis is developing a coding scheme (Bakeman & Gottman, 1986). In this case, each member in the special interest community is assigned an unique U_ID (list user identification) which is associated with personal

EMAIL ADDRESS and NAME. Two useful counter fields are: POST COUNT, the number postings the list user contributed; and REPLY COUNT, the number of public replies (through postings) the user received. The individual data can be arranged in 5 columns as depicted in a *list user table* (Table 8).

| | | | | |
|---|---|---|---|---|
| U_ID | EMAIL ADDRESS | NAME | POST COUNT | REPLY COUNT |

**Table 8 Header Fields of List User Table**

Messages posted to a list server, usually consist of some topical thread, we can assign an unique S_ID to messages with a same SUBJECT LINE together. The SUBJECT COUNT field can record the number of messages with the same topical thread. The data can be arranged in the following 3 columns *topical thread table* (Table 9).

| | | |
|---|---|---|
| S_ID | SUBJECT LINE | SUBJECT COUNT |

**Table 9 Topic Thread Table**

The primary data type for listserver analysis is an email message posted to the list. We can consider each individual posting to the list with the following key fields: each posted message is assigned with an unique MSG_ID; the date of a posting is transformed into equivalence in MDT (Mountain Daylight-saving Time) for uniform standard; the SEC field is the time format which UNIX used internally (based on the elapse seconds since the first second of 1977); and from the subject line of a message, we can obtain the S_ID from the topical thread table (Table 9). We can also extract a message sender's email address and name, then determine the sender's U_ID as FROM: U_ID field; similarly if the massage posted was in response to a prior posting, we can determine to the cited message's MSG_ID and from the cited message, we can determine its poster's U_ID from email address and list it to REPLY TO: R_ID field. The cited message's MSG_ID is then listed in the REPLY TO: MSG_ID field;  Finally, each posting can be assigned a SYMLOG adjective rating scale as SYMLOG CODE (the SYMLOG methodology will be explained in later sections). The data for each posted message can be arranged in 8 columns as depicted in a *message table* (Table 10).

| MSG_ID | DATE | SEC | S_ID | FROM: U_ID | REPLY TO: R_ID | REPLY TO: MSG_ID | SYMLOG CODE |
|---|---|---|---|---|---|---|---|

**Table 10 Header Fields of Message Table**

## *7.3 Preliminary Data Preparation*

We need to prepare the data systematically for extraction of key fields within an e-mail posting. For convenience, we will use the standard UNIX mailbox format which allows us to convert a listserver archive into HTML documents through Hypermail (EIT, 1994). Messages in the UNIX mailbox format archives are typically RFC 822 mail messages appended to each other that look similar to this:

```
From john@foo.com  Mon Jan  1 00:01:30 1994
Date: Mon, 1 Jan 1994 00:01:15 PDT
From: john@foo.com
To: everyone@foo.com
Subject: Hello, world!

Hi, everyone, just saying hello!

From someone.else@foo.com  Mon Jan  1 00:02:00 1994
Date: Mon, 1 Jan 1994 00:01:45 PDT
...
```

The messages are typically separated by lines in this format:

```
From www-talk@www0.cern.ch  Fri Jul  1 00:18:20 1994
```

*Hypermail* (EIT, 94) is a program that takes a file of mail messages in UNIX mailbox format and generates a set of cross-referenced HTML documents. Each file that is created represents a separate message in the mail archive and contains links to other articles, so that the entire archive can be browsed in a number of ways by following links. Archives generated by Hypermail can be incrementally updated, and Hypermail is set by default to only update archives when changes are detected.

Each HTML file that is generated for a message contains (where applicable):

- the subject of the article,
- the name and email address of the sender,

- the date the article was sent,
- links to the next and previous messages in the archive,
- a link to the message the article is in reply to, and
- a link to the message next in the current thread.

After we convert the list into HTML documents with the appropriate fields of our coding scheme already organized systematically, we can start to encode the fields. For content-free analysis purpose, once we have encoded every code fields (expect SYMLOG CODE) for posted messages, we can start to generate two key measurements: time-series plot of the message flow and sociogram of the social network.

## *7.4 Time-series Graph and Analysis*

Time-series analysis offers a wide range of analytic options, and further more, it is possible to create time-series data from the streams of codes as organized in a message table (Table 10). Bakeman and Gottman (1987) have given a comprehensive introduction to time-series analysis. In this section, the method for constructing time-series of message flow will be discussed.

### 7.4.1 Time-series Plot

One advantage of creating time-series graph (Figure 15) from a stream of posted messages is that we can obtain an overall visual picture of the interaction in the listserver.

First, we should determine the time range for the plot, by noting the dates of the first and the last messages in the list archive. Then we determine the number of intervals we want in the time-series plot. From which, we can determine the length of the intervals by a simple calculation:

(1) first we normalize the date of each message based on the standard UNIX elapse time;

(2) then we compute interval = (last message date – 1st message date + 1) / number of intervals.

Afterward, we iterate through the list and tally up number of postings per day. Finally, we create a daily message frequency table: with information such as total number of days; totally number of messages; and number of days with message postings (Table 11).

**Total No. of Days = 316; No. of Messages = 111; Days with Postings = 50**

**…**

| DAY # | DATE | Number of Postings | As Percentage of Total Postings |
|---|---|---|---|
| **D193** | 97/3/7 | 1 | 0.9% |
| **D214** | 97/3/28 | 1 | 0.9% |
| **D273** | 97/5/26 | 1 | 0.9% |
| **D274** | 97/5/27 | 1 | 0.9% |
| **D275** | 97/5/28 | 4 | 3.6% |
| **D276** | 97/5/29 | 6 | 5.4% |

**…**

**Table 11 Partial Listing of a Daily Message Frequency Table**

From a daily message frequency table, we can create a list that records the number of postings per time-series plot interval; for example in Table 11, if we have an interval length of 3 days, the interval D273-D275 then contains 1+1+4 = 6 messages.

Finally, a time-series graph of message traffic (Figure 15) can be plotted, where Y-axis represents the number of postings per interval (X-axis). The time-series plot allows us to visualize the vitality of a list server with respect to the time continuum. In the next chapter, we will see how the graph is used to analyze a particular special interest community as a sample study case. In the next subsection, we will examine statistics issues in time-series analysis.

### 7.4.2 Autocorrelation Statistics

A major problem with analyzing data when they are collected along the time continuum is the inability of conventional inferential statistics to yield reliable results. The crux of the problem is that of human interactions are usually serially dependent along the time continuum (Suen & Ary, 1989). Once the time-series plot were available, we can controlled for autocorrelated time series, as suggested by several authors (Bakeman & Gottman, 1987; Suen & Ary, 1989; Losada, Sanchez, & Noble, 1990).

The autocorrelation function (ACF) is fundamentally a set of Pearson's product-moment correlation coefficients (McClave & Dietrich, 1991). The temporal distance between two observational points in time is referred to as the *lag*. Lag-1 ACF occurs if two observations were made at consecutive points in time.

The general formula for the computation of a lag-k ACF is:

$$\text{Lag-k ACF} = \frac{\sum\left(X_t - \overline{X}\right)\left(X_{t+k} - \overline{X}\right)}{\sum\left(X_t - \overline{X}\right)^2}$$

**Equation 3 Computation of a lag-k ACF**

Where $X_t$ is the observation made at made t, $\overline{X}$ is the mean of all observations, $X_{t+k}$ is the observation made at time t + k, and k is the lag width. There are a number of methods for test statistical significance of an autocorrelation. The most common method is the Bartlett test (Suen & Ary, 1989). With the Bartlett, the standard deviation of the theoretical sampling (i.e., the standard error) of a lag-k ACF is given by:

$$\text{SE lag-k ACF} = \sqrt{1/N(1+2\sum_{i=1}^{k} ACF_i^2)}$$

**Equation 4 Standard Error of a Lag-k ACF**

N is the number of pairs of observations for the ACF, and $ACF_i$ is the lag-i ACF. Therefore, for an approximately .05 level of significance, one can compare the observed value of the lag-k ACF against ± 2 (SE). If the observed value of the lag-k is greater in absolute value, then the null hypothesis of ACF = 0 can be rejected. If the observed ACF is within the range of ± 2 (SE), one fails to reject the null hypothesis of no autocorrelation.

The Auto Regressive Integrated Moving Average (ARIMA) approach is a set of powerful statistical techniques through which the exact nature of the serial dependency of a set of observations made a period of time can be assessed. The details of the ARIMA approach are beyond the scope the present dissertation and can be found in texts by Suen and Ary (1989) and Brown and Rothery (1993).

## *7.5 Social Network Analysis: Sociomatrix and Sociogram*

Once we have an idea of how the traffic flows in a listserver from examining its time-series plot, we can start to investigate its social structure by studying the sociomatrix and sociogram associated with the special interest community. This section introduces sociometric and sociomatrix notions for measuring effective relations between people (Wasserman & Faust, 1994).

One of the primary uses of social network analysis is the identification of the "most important" members in a social network. In this section, we will examine a variety of measures designed to highlight the differences between influential and non-influential members. Definitions of *importance*, *influential* or synonymously, *prominence*, have been offered by many writers (Wellman & Berkowitz, 1988; White & McCann, 1988; Wasserman & Faust, 1994). All such measures attempt to describe and measure properties of "member location" in a social network. Members who are the most important of the most prominent are usually located in strategic locations within the network. Researchers also have attempted to quantify the notions of sociometric "stars" and "isolates" (Wasserman & Faust, 1994).

A social network data table is consisted of measurement on a variety of relations on one or more sets of members. The next subsection describes how to construct a sociomatrix table.

### 7.5.1 Sociomatrix

First, we iterate through a list archive, using a message table (Table 10) and a user table (Table 8) as references, create a *sociomatrix* (Table 7). A sociomatrix is of size g × g (g rows and g columns), where g is the total number of nodes in a graph, in our case, g denotes the total number of members in the list. It is sometimes referred to as a *valued adjacency* matrix by graph theorists (Wasserman & Faust, 1994). The entries in a sociomatrix, $X_{ij}$, record which pairs of nodes are adjacent and if adjacent by how much. In the case of message flow, it indicates how many postings member i **publicly** replied to member j. If the value is greater than 0, members i and j are considered as adjacent nodes in a sociogram. The resulting sociomatrix allows us to calculate useful measurements of a social network, such as: total numbers of messages posted and referenced, nodal degrees, density, member degree centrality, and member degree prestige (described in later subsections).

As depicted in the punctuated discourse model (Chapter 5), a response to a list can be: (1) specially targeted to the community as a whole; (2) purposely replied to a specific individual and to the community altogether; and (3) intentionally addressed to a specific individual with public discourse as a side-effect. However the second and third intentions exhibit the same overt behaviour, that is, posting a message to another member publicly via a list.

### 7.5.2 Sociogram

From a sociomatrix, we can draw a *directed-graph* sociogram (Figure 16), which allows us to visualize the social network among members in the list with respect to their posting patterns. A node represents a member in the special interest community. Each value on a directed arc between two nodes represents the number of messages member i posted in respond to member j. For example, in Figure 16, member $M_2$ had posted 5 messages in *public* responses to member $M_3$. As mentioned earlier, a sociogram allows us to visually identify who has greater influence than others in the social network. Similarly, the degree centrality measurements from a sociomatrix table also allows us to reach a similar conclusion.

### 7.5.3 Messages Posted and Referenced

From a sociomatrix table, the easy kinds of measurements we can compute are the numbers of messages each member posted and referenced publicly with *respect to one another*. If have we have a sociomatrix X, the *total number of messages posted to other members by each member* $n_i$ at row i is the sum of all column entries in the row i:

$$t_O(n_i) = \sum_{j=1}^{g} X_{ij}$$

**Equation 5 Total Number of Posts to Others**

Similarly, the *total number of responses which publicly referenced prior messages posted by member* $n_i$ is the sum of all row entries in the column:

$$t_R(n_i) = \sum_{j=1}^{g} X_{ji}$$

**Equation 6 Total Number of References Received from Others**

Note the first measure $t_O(n_i)$—the total number of postings member $n_i$ made with respect to other members—may be different than the total number of postings member $n_i$ made to the list, $t_L(n_i)$. This is due to the fact the members often post messages on the list as *simple undirected broadcasts*. Those messages are *intensional* in nature without particular *extensional* awareness. Therefore, $t_L(n_i) \geq t_O(n_i)$ is always true.

Since a sociomatrix only contains entries members made to each other, we will need a separate set of frequency counters to keep count of $t_L(n_i)$.

## 7.5.4 Nodal Degrees and Network Density

A sociomatrix table also allows us to calculate the nodal degrees of each member. The **indegree** of a node is the number of nodes incident *to* the node (the number of arcs terminating at it) and the **outdegree** of a node is the number of arcs incident *from* the note (the number of arcs originating from it). Notice that row i of a sociomatrix contains entries $X_{ij} > 0$ if node $n_j$ is incident from node i. The number of columns with value > 0 is thus the number of nodes incident from node $n_j$ and is equal to the outdegree of node $n_i$. Similar the entries in column i of a sociomatrix contains entries $X_{ji} > 0$ if node $n_i$ is incident to node $n_i$. Thus the number of row with value > 0 is equal to indegree of node $n_i$. They are formally defined in the following equations.

$$d_O(n_i) = \sum_{j=1}^{g} X'_{ij} \text{ where } X'_{ij} = 1, \text{ if } X_{ij} > 0$$

**Equation 7 Calculation for Outdegree**

and

$$d_I(n_i) = \sum_{j=1}^{g} X'_{ji} \text{ where } X'_{ji} = 1, \text{ if } X_{ji} > 0$$

**Equation 8 Calculation for Indegree**

The density of a digraph (directed graph) can be calculated as the sum of all entries in the normalized matrix, divided by the possible number of entries. In this case, a normalized matrix entry means that every value in the original matrix greater than 0 will be transformed into 1 in the normalized matrix. That is, $X'_{ij} = 1$ if $X_{ij} > 0$.

$$\Delta = \frac{\sum_{i=1}^{g}\left(\sum_{j=1}^{g} X'_{ij}\right)}{g(g-1)} \text{ where } X'_{ij} = 1, \text{ if } X_{ij} > 0$$

**Equation 9 Computing Network Density**

### 7.5.5 Member Degree Centrality

The simplest definition of member centrality is that central members must be the most active in the sense that they have the most ties to other members in the network. No where is this easier seen than by comparing a graph resembling a star to one resembling a circle. A star graph has the property that exactly one member has ties to all $g - 1$ other members, and the remaining members have only their single tie to the first member. A centre member (i.e., the central node) is clearly the most active, and one could view this high level of activity as a large amount of centrality. This very active member should, thus, have the maximal centrality index. Here we measure centrality simply as degree (Wasserman & Faust, 1994). A circle graph has no member more than active than any other member, so all members should have exactly the same centrality index.

The outdegree of a member is important in determining her influence; therefore, a centrality measure for an individual member should be the outdegree of the node, $d_O(n_i)$. $C_D(n_i) = d_O(n_i)$. We notice that one problem with this measure is that it depends on the group size g; indeed, its maximum value is $g - 1$. Consequently a proposed standardization of the measure is the proportion of nodes that are adjacent to $n_i$.

$$C'_D(n_i) = d_O(n_i) / (g - 1)$$

**Equation 10 Member Degree Centrality**

$C'_D(n_i)$ is independent of g, and thus can be compared across networks of different sizes.

### 7.5.6 Member Degree Prestige

With directional relations, responses a member received publicly are quite interesting to a network analysis. Thus measures of centrality may not be as much a concern as measures of prestige. Both centrality and prestige measures should be computed for digraph, since they do attempt to measure different structural properties.

The simplest member level measure of prestige is the indegree of each member, which we denoted by $d_I(n_i)$ in Equation 8. The idea is that members who are prestigious tend to have many respondents referenced their prior postings. So we define: $P_D(n_i) = d_I(n_i)$. As with the comparable degree centrality measurement based on outdegree, it is dependent upon group size g; thus the standardization.

$$P'_D(n_i) = d_I(n_i) / (g - 1)$$

**Equation 11 Member Degree Prestige**

$P'_D(n_i)$ gives us the proportion of members who referenced prior postings from member $n_i$. It is sometimes called a "relative indegree" (Wasserman & Faust, 1994). The larger this index is, the more prestigious is the actor. Maximum prestige occurs when $P'_D(n_i) = 1$; that is when member $n_i$ has been referenced by all other members.

The centrality index discussed in the last subsection is originally designed for a graph (and thus, symmetric sociomatrices); however, it can also be modified for use by a digraph. The notion of "prestige", however, can only be quantified by using relations for which we can distinguish "postings" sent from postings received by the members, and therefore, can only be studied with a digraph. With directional relations, measurements such as outdegree and indegree are likely to be different, and prestigious members are usually those with large indegrees, or postings referenced.

A cautionary note is needed here. The notion of "prestige" is disembodied from the external reality, since we are not concerned with many social attributes of a list member in real life (e.g., social standing; popularity). We can only obtain measurements solely based on overt behaviours exhibited by members. Therefore, a better term for "prestige"—in the cyberorganism context—may be that of "*social power*" (as described in Chapter 5). *Prestige*, in this context, is the ability of an individual to invoke and to generate responses from other members in a 'virtual' social network. A member who has a high *member degree prestige* may not be truly 'prestigious' in real life; but in the context of the special interest community delineated by a listserver, he or she has the prestigious 'social power' to influence others' behaviours. In this dissertation, we will retain the use of "prestige" in the sense of the social network analysis terminology (i.e., *member degree prestige*).

Both centrality and prestige indices are examples of measures of the prominence or importance of members in a social network. In Chapter 8, we will see a sample case of how those measures can be used in analyzing the social network in a special interest community.

## *7.6 SYMLOG*

The above sections describe measurements which can be calculated without content-analysis. This section, I will examine a methodology dealing with analysis of group-dynamics based on content-analysis of list archives.

### 7.6.1 Theoretical Background

SYMLOG, an acronym for the Systematic Multiple Level Observation of Group is a 'new' *field theory* for the analysis of group dynamics that adds systematic measurement techniques to a perspective proposed by Lewin (1951). The theory, developed by Bales and his associates (Bales & Cohen, 1979) assumes that interaction can be measured in terms of three orthogonal dimensions: **Up-Down** (U-D dimension) whose psychological correlate is represented by dominant vs. submissive behaviour; **Positive-Negative** (P-N dimension), psychologically mapped into friendly vs. unfriendly; and **Forward-Backward** (F-B dimension), whose psychological correlate is task-oriented vs. emotionally expressive behaviour. The image of an individual in this space can be located using a 26 item check list or by direct coding of interaction.

There are two general classes of communication behaviour in group-dynamics. The first is socio-emotional behaviour, represented by positive and negative actions like seemingly friendly, showing tension, and dramatizing, the second is task behaviour, represented by suggestions, opinions, and information. In investigating leadership, Bales has found that typically the same social group will have two different kinds of leaders who are the decider critical subsystems in a team or special interest community. A *task leader*, who facilitates and coordinates the task-related comments, directs energy toward getting the job done. The emergent of the task leadership role is essential for the problem-solving activity in the group. Equally important is the emergent of a *socio-emotional leader*, who works for improved relationship in the group, concentrating on interactions in the positive and negative sectors. Usually, the task and socio-emotional leaders are different people (Littlejohn, 1992).

### 7.6.2 SYMLOG Coding Scheme

SYMLOG uses these 3 dimensions—UD, PN, FB—to code interaction at two levels: the behavioural level and the image level (which addresses the meaning conveyed by behavioural acts and consequently, is more interpretative than descriptive). The behavioural level comprises both overt and non-verbal behaviour.

Usually a SYMLOG coder must enter by hand the following information: (a) the time of the event (approximated to time of a message posting); (b) who is the sender; (c) who is the received of the action; (d) a specification of whether the observed behaviour was overt or nonverbal; (e) the behavioural code; (f) a comment describing the behaviour topic (Losada & Markovitch, 1990).

As stated earlier in this chapter, we will only focus on the overt behavioural level within the cyberorganism research; hence, our SYMLOG coding scheme will be simplified to use as the following:

(a) Message MSG_ID associated with DATE & TIME fields

(b) SENDER U_ID

(c) RECEIVER R_ID

(d) ‘*Verbal*’

(e) SYMLOG CODE

(f) SUBJECT S_ID

Since the field (d) will always be the same as '*Verbal*' for all messages, it can be omitted. Therefore, the rest of the coding fields can be merged into the message table (Table 10). The (e) SYMLOG CODE is the predominate behaviour observed in a message posting; It is rated using the *SYMLOG adjective rating list* (Bales & Cohen, 1979). The list is depicted in Table 12.

| | |
|---|---|
| U: | active, dominant, talks a lot |
| UP: | extroverted, outgoing, positive |
| UPF: | a purposeful democratic task leader |
| UF: | an assertive, business-like manager |
| UNF: | authoritarian, controlling, disapproving |
| UN: | domineering, tough-minded, powerful |
| UNB: | provocative, tough-minded, powerful |
| UB: | jokes around, expressive, dramatic |
| UPB: | entertaining, sociable, smiling, warm |
| P: | friendly, equalitarian |
| PF: | works cooperatively with others |
| F: | analytical, task-oriented, problem-solving |
| NF: | legalistic, has to be right |
| N: | unfriendly, negative |
| NB: | irritable, cynical, won't cooperate |
| B: | shows feelings and emotions |
| PB: | affectionate, likable, fun to be with |
| DP: | looks up to others, appreciative, trustful |
| DPF: | gentle, willing to accept responsibility |
| DF: | obedient, works submissively |
| DNF: | self-punishing, works too hard |
| DN: | depressed, sad, resentful, rejecting |
| DNB: | alienated, quits, withdraws |
| DB: | afraid to try, doubts own ability |
| DPB: | quietly happy just be with others |
| D: | passive, introverted, says little |

**Table 12 SYMLOG Adjective Rating List**

After every message has been rated with a SYMLOG CODE as seen in adjective rating list, we can calculate the raw interaction scores for each member on the list. The raw interaction scores for each member are the frequencies of incident ('*act'*) with respect to *U, D, P, N, F, B* directions depicted in each rated SYMLOG CODE. For example, if a message sent by member U_21 had been rated as *DN:* we then add 1 *D* and 1 *N* to $D_{21}$ and $N_{21}$ counters respectively. Also, there should be one entire-group frequency counter for each one of the *U*, *D*, *P*, *N*, *F*, *B* directions. Those counters will assist in our calculations for aggregations of SYMLOG interaction scores.

### 7.6.3 Aggregating SYMLOG Interaction Scores

The following aggregating procedure will be discussed in terms of P-N dimension, but it is designed to apply to other two dimensions as well (i.e., U-D and F-B).

Once the raw SYMLOG Interaction Scores have been computed for all members, we can aggregate the SYMLOG coded messages for the entire listserver to obtain the following:

$p_i$ = number of acts by the $i^{th}$ member scored P

$n_i$ = number of acts by the $i^{th}$ member scored N

$P$ = total number of acts by scored P for the entire group

$N$ = total number of acts scored N for the entire group

$m$ = number of members

The formula for transforming the raw scores into aggregated data for plotting SYMLOG Field Diagram can represented as Equation 12 (Bales & Cohen, 1979). The resulting F value usually should be within the range of -18 and 18.

$$F(p_i, n_i, P, N, m) = C(m) \cdot \left( m\sqrt{\frac{p_i^2 + n_i^2}{P^2 + N^2}} + 1 \right) \cdot Q(p_i, n_i)$$

$$\text{where } C(m) = 18 / \left( 1 + \frac{m}{\log_e(m+5) + .58} \right)$$

$$Q(p_i, n_i) = \frac{p_i - n_i}{p_i + 0.4n_i} \text{ if } p_i \geq n_i$$

$$Q(p_i, n_i) = \frac{p_i - n_i}{n_i + 0.4p_i} \text{ if } n_i > p_i$$

**Equation 12 Formula for Calculating Field Location**

An *Original Field Diagram* (Figure 17) drawn from the aggregating interaction scores will fall within our desired limit of 18 in many cases but unfortunately not in all, because of the differences in the group size and the fact in some groups the participation is very unequal. Hence, one more scaling operation is regularly formed for the purpose of plotting the constellation of points on the Original Field Diagram. This step is the adjustment of the regular plotting scores by the method of '*Expansion Multiplier*' (Bales & Cohen, 1979).

It is actually a factor that will *expand* the constellation of points when they are too closely clustered to fill the diagram, but will *contract* the constellation of points when any of them exceeds the scale limit of 18. In addition it takes into account of the size of the circles one will use in plotting the U-D position. For this reason, the Expansion Multiplier cannot be found until all three dimensions have been computed for all individual members.

First, we need to plot the original diagram then determine the outermost field circle and its Circle Radius. We then calculate the Expansion Multiplier (sign is always +) as:

$$\frac{18\ -\ \text{UD Circle Radius}}{\text{Original Center Point Location}} = \left[ExpansionMutiplier\right]$$

(change - sign to +)

**Equation 13 Expansion Multiplier Formula**

Once the proper Expansion Multiplier is found, it is applied to the location of all points in both P-N and F-B dimensions. The modified points then can be plotted in an Adjusted Field Diagram.

In addition, if there are more than one coder involved in rating SYMLOG Interaction Scores (either for efficiency or reliability reasons), we can aggregate their summary scores systematically. Bales and Cohen's SYMLOG manual (1979) provides two methods to calculate estimates for *inter-coder agreement* and *scoring reliability estimate*. Similarly, Bakeman and Gottman (1986) have introduced similar statistics (e.g., Cohen's kappa) to address observer agreement issues.

### 7.6.4 Interpretation of SYMLOG Field Diagram

The main output from applying the SYMLOG coding procedures is a *Group Average Field Diagram* which summarized the average group behaviours by each member as a circle whose radius conveys the level of dominance. The larger the circle, the more dominant the person. Thus, the group relationship can be projected onto a 2 dimensional Field Diagram (Figure 17): with the P-N and F-B dimensions as the horizontal and vertical axes and the U-D dimension represented by circle size.

Bales argues this graphical representation of group-dynamics is more powerful than numerical information: "*the visual diagrams seem to allow […] observations of group processes and group relationships in a concrete and observable manner and to speak about them much more easily and directly than they generally can about numerical data*" (Bales & Cohen, 1979).

When the images of members of a social network are located on a Field Diagram, it is possible to identify the extent to which the community may be unified or polarized. If a community is polarized, the theory predicts which individuals may become *mediators* or *scapegoats* (Hare, 1989). Predictions can also be made about the parts played by dominant members, isolates, and individuals who appear to occupy other locations in the three dimensional space.

## *7.7 Summary*

Special interest communities are associator subsystems situated in an integrated architecture for communication, social, and knowledge processes. Mailing list servers handle the primary discourse of the special interest communities. This chapter presents systematic methodologies for analyzing listservers

In order to analyze the social network of a special interest community, we first obtain three types of measurements from list members' interactions: time-series plot, sociogram, and SYMLOG field diagram. The chapter introduces an overall coding scheme with key fields, such as: who posts to the list, who replies to whom, or which topic thread is growing. The coding scheme is used to produce various coding tables. Many useful descriptive statistics then can be determined from the resulting coding tables.

For content-free analysis, time-series plots can be drawn from the daily message frequency tables. From those basic statistics, sociomatrices, sociograms or directed-graphs (Wasserman & Faust, 1994) of the list can be determined and further analyzed. Measurements like indegree, outdegree, degree centrality, and degree prestige can be calculated for identifying important or influential members in the social network of a list. For content analysis, SYMLOG field diagrams (a system for the multiple level observation of groups, by Bales & Cohen, 1979) can be plotted after sequential event-coding of posted messages. Together sociograms and SYMLOG field diagrams allow us to visualize the group-dynamics of special interest communities.

To conclude, this chapter introduces systematic methodologies for analyzing social networks and dynamics of special interest communities. The next chapter presents *ListA*, a *Listserver Analyzer* program, for automating coding, tabling, and graphing procedures in the analysis methodologies. A special interest community will be selected as a sample case for demonstrating the methodologies and the ListA program.

# CHAPTER 8

# ListA: Listserver Analyzer

This chapter describes the *ListA: listserver analysis* program. It applies the analysis methodologies described in Chapter 7 and utilizes the ListA program to a sample case listserver as a demonstration.

## *8.1 Software Tools for Listserver Analysis*

For the research purpose of analyzing special interest communities, I had developed some software tools to assist data collection and analysis procedures as described in the last chapter.

Various Perl scripts were written to be used for extracting key data fields from listserver archives and to prepare them for the construction of time-series plots, sociograms, and descriptive statistics. *ListAnalyzer (ListA)*[5], a web-based program, was designed to unify these software tools as an interactive package that can be accessed via the Netscape Navigator™ browser (version 3.0 or later).

These list analyzer tools offer great cross-platform portability. As a result, these automation tools also allow researchers in remote sites to apply integrated methodologies in a systematic manner. In addition, ListAnalyzer's utilization of *client-server partition* (Shaw & Gaines, 1996) allows distributed researchers to cooperate in their list analyses. The client-side JavaScript 1.1 enabled browsers (e.g., Netscape Navigator™ Navigator 3.0 or Microsoft Internet Explorer™ 3.0) can be used for the data encoding anywhere researchers have access to the web; and the server-side Perl CGI database and analysis

---

[5] The program can be accessed via http://ksi.cpsc.ucalgary.ca:8800/cgi-bin/lee/listA.pl

engine can reside in a centralized location for comparative analyses of different listservers done by various researchers.

## *8.2 System Overview*

In terms of content-free analysis, ListAnalyzer can automatically produce data files with sorted key fields, a time-series plot and associated descriptive statistics, a sorted sociomatrix, and an interactive sociogram. Together, they can give fast feedback for researchers or/and list members about a special interest community's vitality, social structure, and life cycle. From the resulting feedback, we can determine whether or not further content analysis of the community is worthwhile. In order to analyze socio-psychological group dynamics of the community, we can conduct content analysis using the SYMLOG methodology. ListA facilitates efficient coding of observed interaction following the SYMLOG formalism and generates Group Average Field Diagrams (original and adjusted) once the coding has been completed.

ListA conceptually groups five analysis tools (Table 13) which correspond to procedures described in the proposed systematic methodologies for listserver analysis:

| | |
|---|---|
| PREP | preparation of data files |
| TIME-SERIES | time-series plot |
| SOCIOGRAM | dynamic interactive sociogram |
| ENCODE SYMLOG | encoding SYMLOG interaction scores |
| FIELD DIAGRAM | group average field diagrams (original & adjusted) |

**Table 13 Five Analysis Tools in ListA**

The next section describes operations and user interfaces of the ListA program in details.

### *8.3 Descriptions of ListA Analysis Tools*

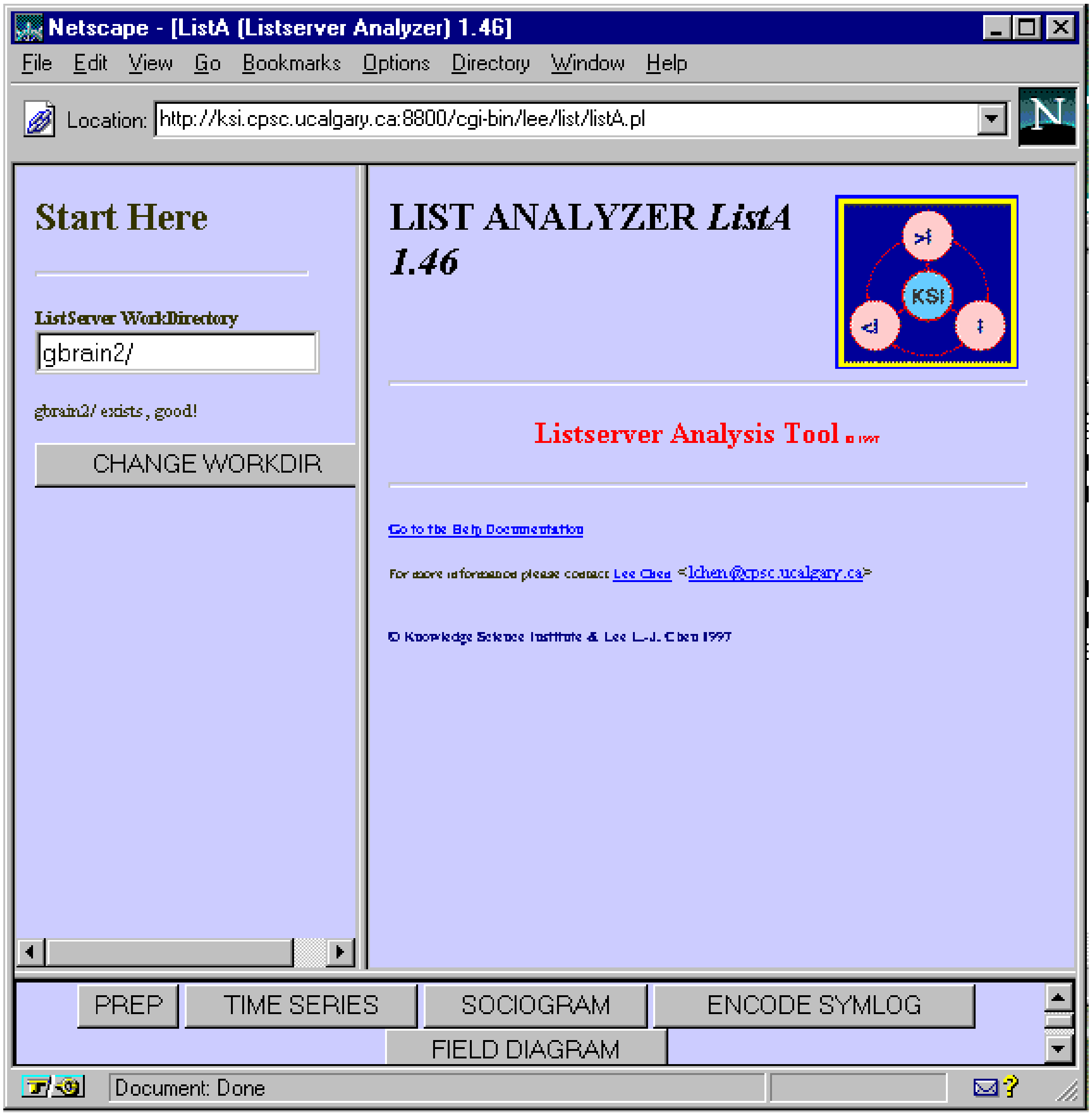


**Figure 18 ListA: Listserver Analyzer Initial Screen Shot**

Before using the ListA program, we need to convert a listserver archive from the UNIX mailbox format into separated HTML documents via Hypermail (EIT, 1994). Once the conversion step is completed, we then collect those HTML documents into a *work directory*. Afterward, we enter the location of the *work directory* into the **Listserver**

**Work Directory** field, then press the **CHANGE WORKDIR** button at the left side pane of the window (Figure 18).

### 8.3.1 PREP: Preparation for Data Files

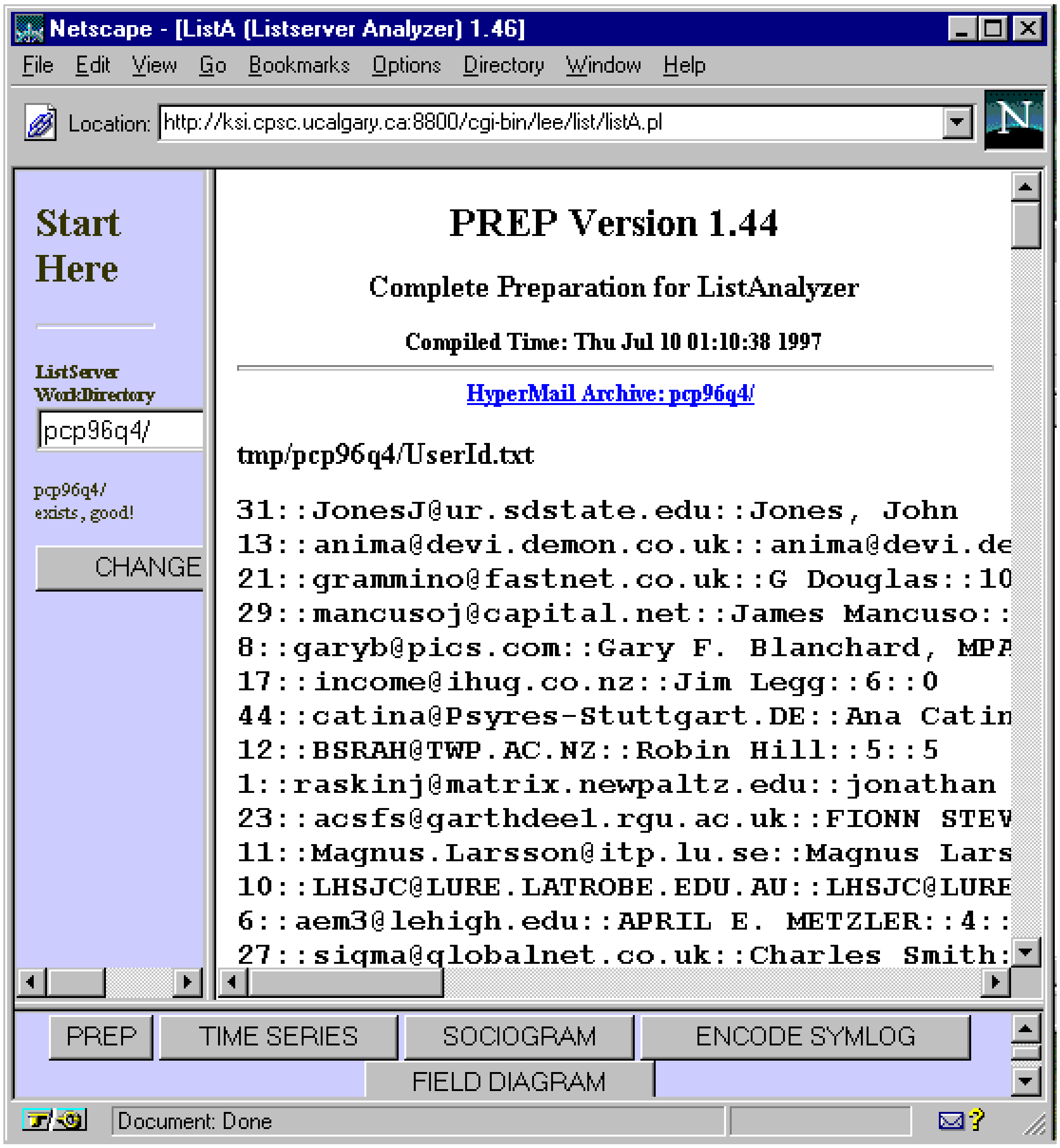


**Figure 19 ListA: Data Preparation**

Once we have changed the current work directory to the one that contains the Hypermail converted HTML messages, we can start the data preparation procedure by pressing the **PREP** button on the *control panel* (at the bottom panel of the window in Figure 19).

During the PREP phase of program operations, ListA extracts key fields from HTML documents then prepares systematic data files containing ranked *User ID List*, *Subject ID List*, *Message Table*, and *Sociomatrix Table* (Figure 19). Each resulting data file has its *corresponding tab-separated-format* file allowing it to export to the Microsoft Excel™ application or other statistical packages like SDIS/GSEQ (Bakeman & Quera, 1995) for further analysis. In addition, during this phase, ListA generates a dynamic sociogram HTML file ready for use in the SOCIOGRAM phase.

## 8.3.2 TIME SERIES: Time-series Plot

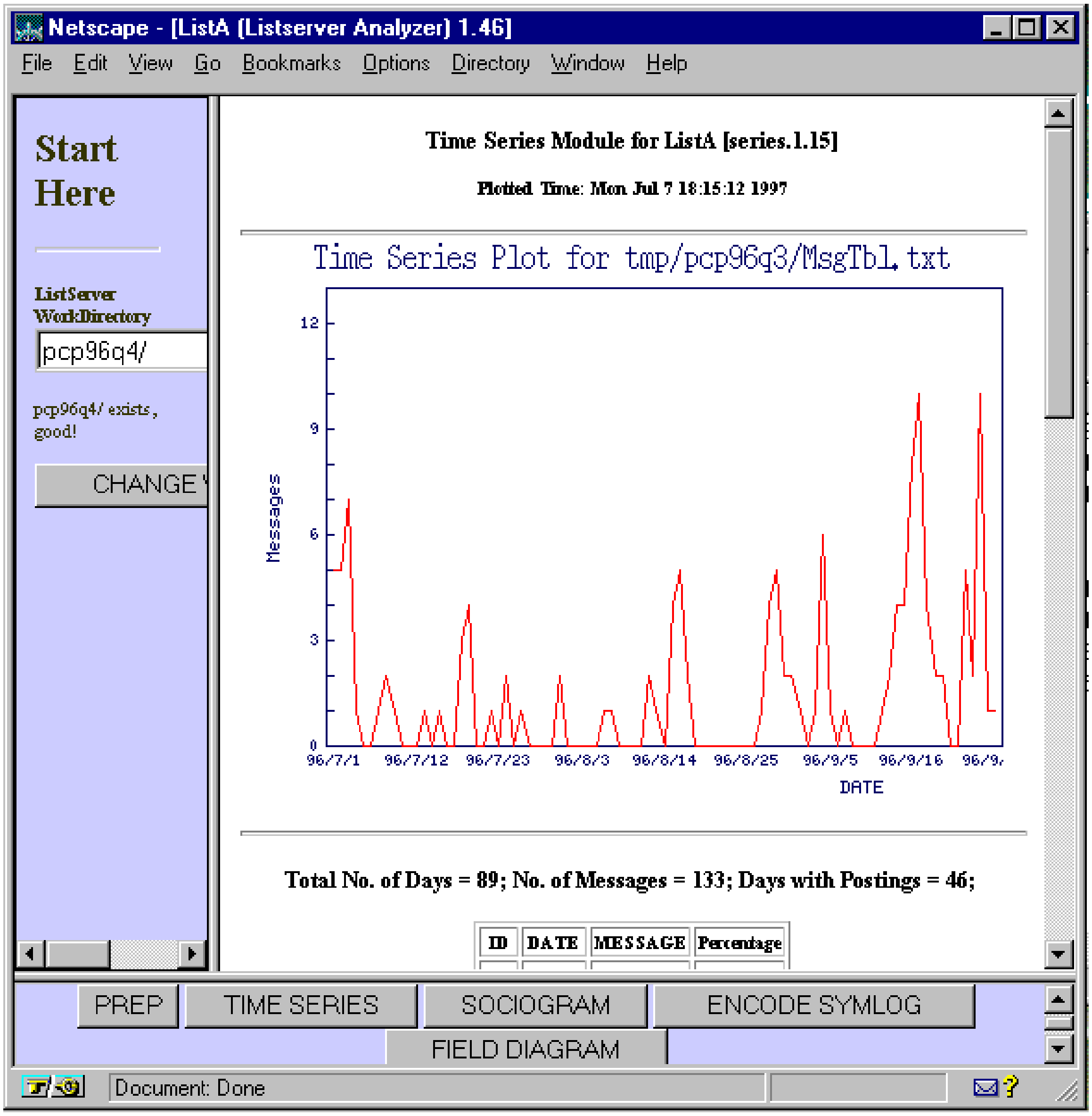


**Figure 20 ListA: Time Series Plot**

After preparing necessary data files during the PREP phase of the program operations, we can obtain a time-series plot and other descriptive statistics by pressing the **TIME SERIES** button at the control panel (Figure 20).

ListA produces a time-series plot of the message traffic flow from the Message Table prepared earlier. In the right pane of the ListA window (Figure 20), we see the resulting time-series plot, together with statistics, such as: total number of days in the period since the first post; number of messages during the period; number of actual days with postings; and a daily message frequency table (Table 11).

## 8.3.3 SOCIOGRAM: Dynamic Interactive Sociogram

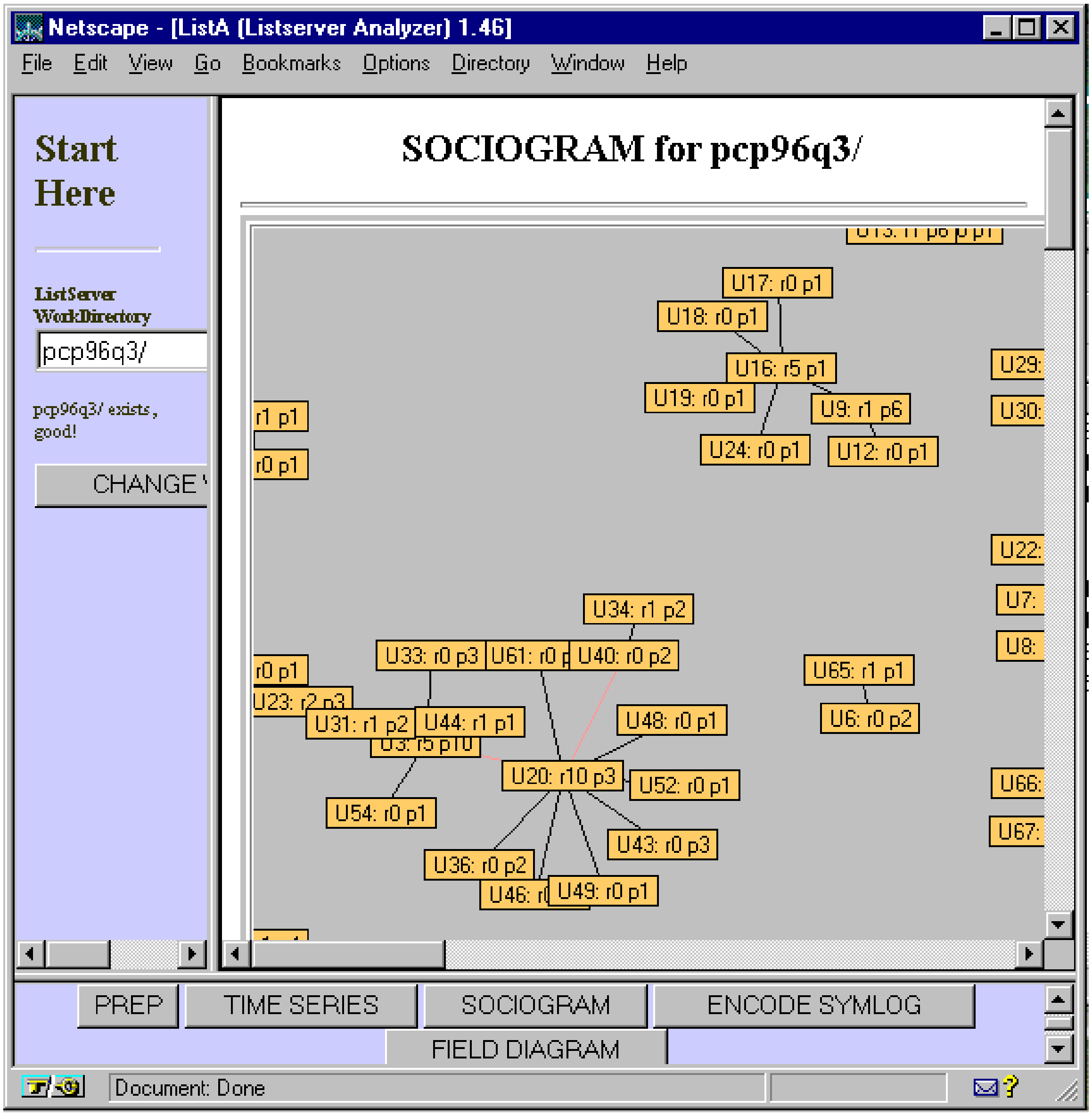


**Figure 21 ListA: Dynamic Sociogram**

During the PREP phase of the program operations, ListA generates a sorted Sociomatrix Data File for the list in the current work directory. We will see parts of an example Sociomatrix Date File generated for our sample listserver later (in Figure 28). The resulting sociomatrix allows us to apply social network analyses and a detailed construct sociogram as described as in Chapter 7. In addition, ListA provides researchers a *dynamic* *interactive sociogram* as an immediate feedback. After pressing the **SOCIOGRAM** button in the control panel, we can interactively manipulate a dynamic sociogram (Figure 21).

A dynamic sociogram is a *Java applet* extended from Sun Micro's *GraphLayout Java Application*. The applet allows researchers to move any node in the sociogram around, other nodes connected to it then follow suit accordingly (based on a *stress-relaxation* algorithm). The stress level for two nodes are defined as the difference between an ideal (predefined) arc length between them and the current arc length in graphical layout. Each pair of nodes try to reduce their stress level by moving closer. Researchers can then rearrange the dynamic sociogram interactively while nodes float around to minimize stresses.

In addition, a pruning algorithm I implemented leaves out members who did not contribute to the list. If one member received any reference from another member, there would be a link (arc) between them. The idea is to enable researchers to be able to visualize the social networks in a list quickly.

Each member node is consisted of 3 fields: $n_i$—*User ID*; $t_R(n_i)$—*total number of referenced received*; and $t_L(n_i)$—*total number of posts made to the list*. For example, a node with the label **U16: r5 p1** (Figure 21) denotes that the list user $n_{16}$ has received 5 publicly cited postings, and has posted 1 message to the list.

In a dynamic sociogram, we can visualize the relative *prestige degrees* of member nodes by noticing the lengths of arcs connected to them. Gradually, a dynamic sociogram settles down into a near stable state (minimum overall stress) in which 'cliques' within a special interest community can be readily identified. For example, in the pcp96q3 list (Figure 21), members U20 and U16 are the most influential numbers in their *social cliques*.

## 8.3.4 ENCODE SYMLOG: Rating SYMLOG Interaction Scores

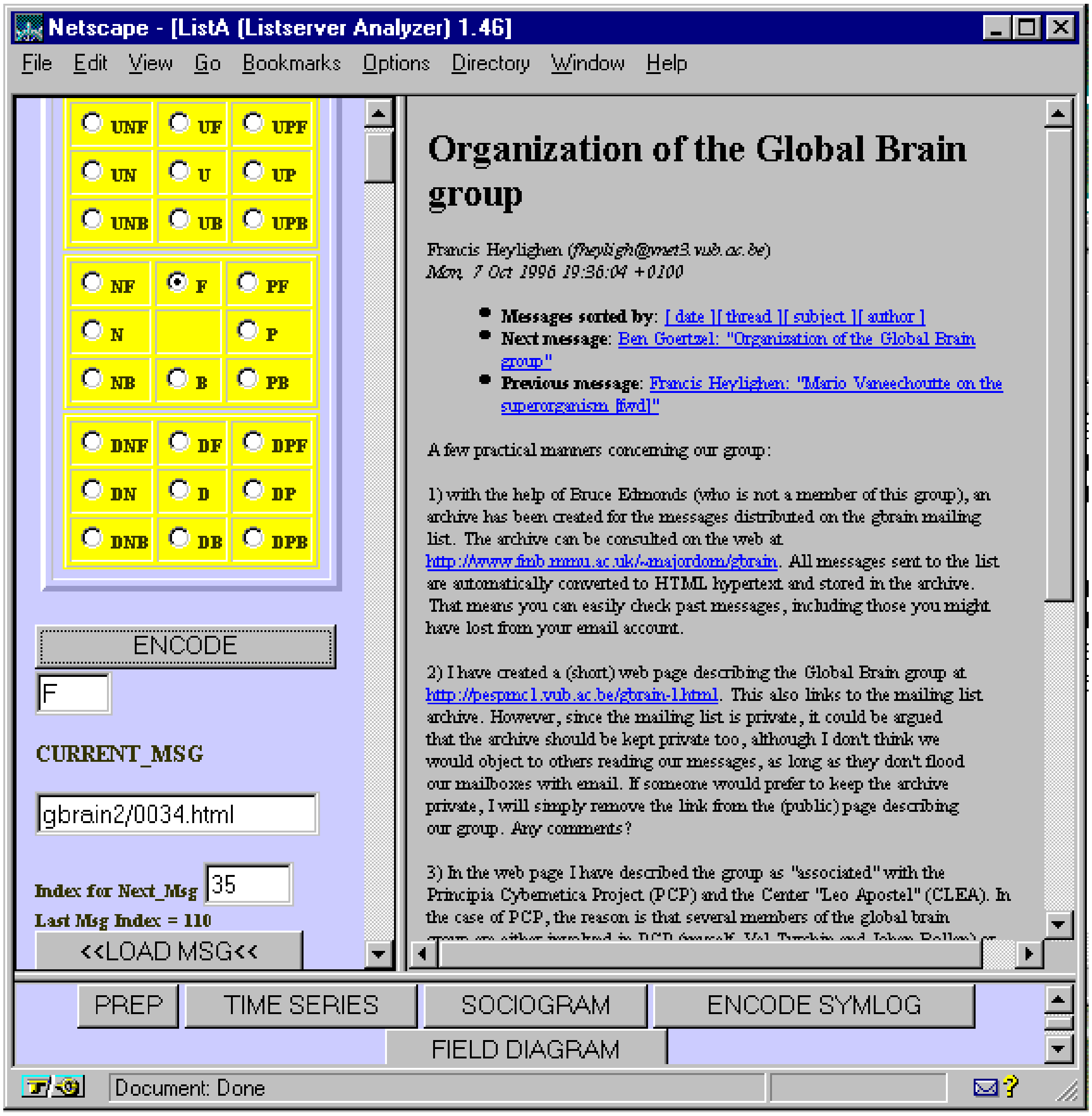


**Figure 22 ListA: SYMLOG Interaction Scores Encoding Phase**

Figure 22 depicts the encoding phase of SYMLOG interaction scores for posts in a list, after pressing the **ENCODE SYMLOG** button on the control panel. The right pane in the window displays a Hypermail generated HTML document presenting a list posting.

ListA first loads into memory a set of SYMLOG data files (each corresponds to one posted message) and determines which messages have not yet been rated. It then creates a *message queue*: an array list containing pointers to HTML posts to be rated. ListA begins by loading the first Hypermail message on the message queue to the right pane for analysis.

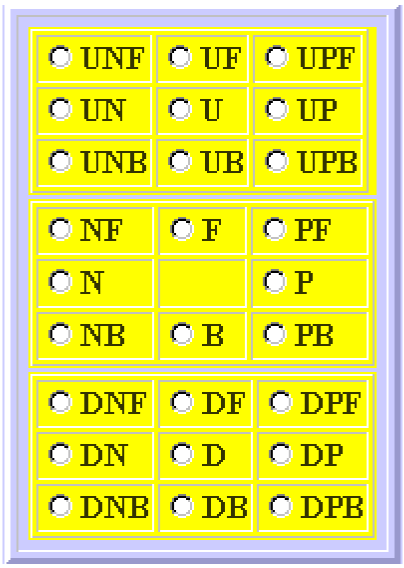


**Figure 23 Radio Buttons for SYMLOG Adjective Rating**

At the top of the left pane, there is a group of 26 radio buttons representing SYMLOG Adjective Rating List (Table 12). After we have analyzed the message content in the right pane, we then select a radio button (Figure 23) corresponding to a SYMLOG adjective rating. By pressing the **ENCODE** button, ListA records the rated interaction score in its temporary database and loads the next message in the message queue to the right pane.

We do not have to follow the message queue in order, sometimes we may want to analyze the list sorted by author or by subject order, instead of by chronology. Because the messages are in Hypermail HTML format, we can jump freely to other messages through hyperlinks on each message, ListA allows us to load any message (displayed on the right pane) for encoding by pressing the **<<LOAD MSG<<** button (on the left pane). Similarly, we can enter a desirable index number in the **Index for Next_Msg** field, then press the **>>NEXT MSG>>** button. ListA then displays the selected message on the right pane for analysis.

We can also examine what we have rated so far by scrolling down the right pane. The temporary database associated by the message queue is arranged in rows of triplet:

(*Message/Index, Corresponding Entry in the Message Table, SYMLOG Code*)

By pressing the **SAVE RESULTS** button, we can transfer the results from the temporary database back to the SYMLOG data files with the newly rated SYMLOG codes. Once every message in the list has been analyzed and rated, ListA displays the following advisory message: "*All Message Have been Rated, Please Run the SYMLOG Field Diagram Phase Now*."

### 8.3.5 FIELD DIAGRAM: SYMLOG Group Average Field Diagrams

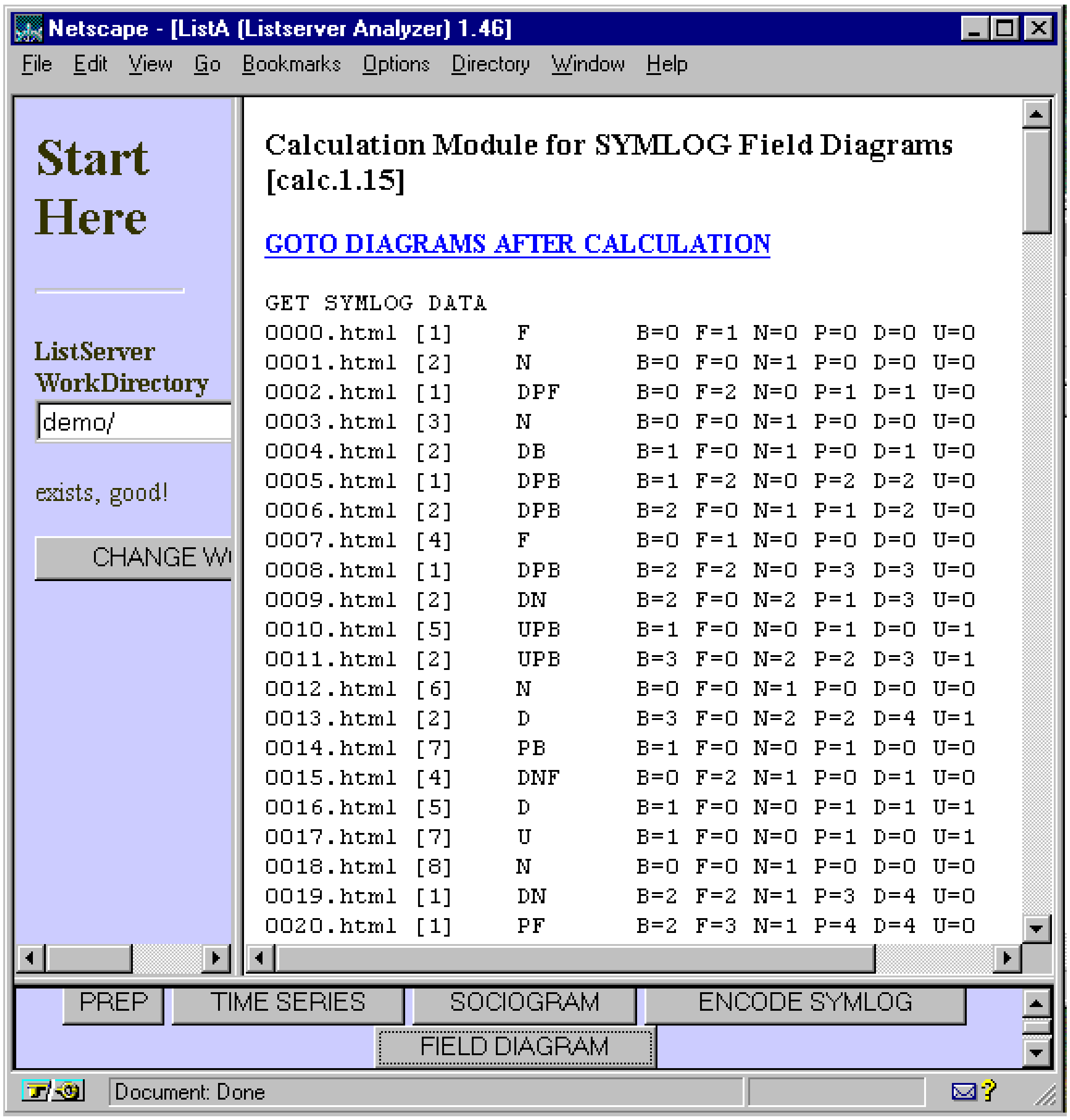


**Figure 24 ListA: Calculation of SYMLOG Interactive Scores**

After we have rated all messages in the list, we can start the SYMLOG analysis phase by pressing the **FIELD DIAGRAM** button on the control panel. During this phase, ListA first calculate members' individual SYMLOG interaction scores in 3 dimensions (U-D, P-N, F-B). Figure 24 depicts the intermediary process of obtaining raw interaction scores for each message 'act'.

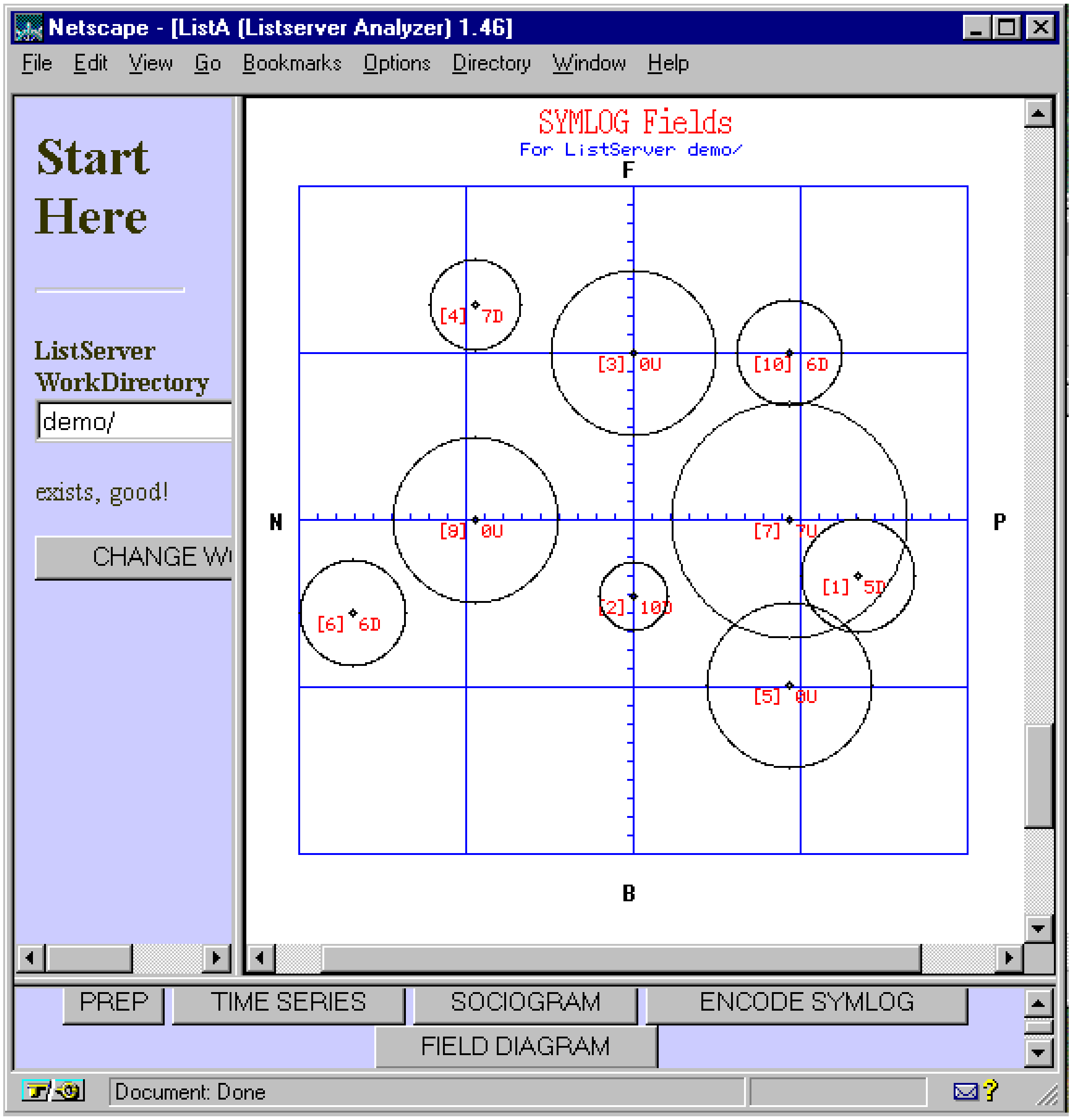


**Figure 25 ListA: SYMLOG Field Diagram (with Adjusted Expansion Multiplier)**

Once the calculations have been completed (based on Equation 12 & Equation 13), we can see two resulting SYMLOG Group Average Field Diagrams. First is the original without adjustment (Figure 17). Second is the one adjusted by applying its *Expansion Multiplier* factor (Figure 25).

Group Average Field Diagrams summarize the average group behaviours by each member as a circle whose radius conveys the level of dominance. The larger the circle, the more dominant the person. Thus the group relationship can be determined with the P-N and F-B dimensions as the horizontal and vertical axes and the U-D dimension represented by circle size.

## *8.4 Sample Case for Listserver Analysis*

In the previous section, we have seen how the ListA program can be used to automate many procedures in the proposed systematic methodologies for listserver analysis. In this section, I will introduce a sample case of special interest community. By analyzing this online community using the ListA program, I will demonstrate how can we apply the methodologies described in Chapter 7.

The **Global Brain** special interest community has been created to discuss "*the emergence of a global brain out of the computer network, which would function as a nervous system for the human super-organism*" (Gbrain, 1996). Its founding character or zygote (i.e., its origin meme) is "*to promotes all theoretical and experimental work that may contribute to the elaboration of global brain theory, including the practical implementation of global brain-like computer systems, and the diffusion of global brain ideas towards a wider public (e.g., by the organization of conferences, or publication of books, on the subject)*" (Gbrain, 1996).

Two main reasons for selecting the global brain (i.e., Gbrain) community as our sample case for analysis are: (a) its relatively manageable size (17 members to date); and (b) my own personal interest with the subject matter. I have been tracking this community since its inception to the present day. It is an ideal candidate for demonstrating the listserver analysis methodologies in details.

The current analysis fully complies with the suggested American Psychology Association (APA) ethical guidelines for studying cyberspace communities (Storm King, 1996). Internet communities exist with varying degrees of perceived privacy. This factor represents the degree to which group members perceive their messages to be private to that group. In our case, the issue of "when is it ethical to study a special interest community without the informed consent of its participants?" is mitigated by the factor of a low degree of perceived privacy. All participants on the Gbrain listserver are aware of the extreme public nature of their discourse; in fact, they have set up, on the web, a mirroring Hypermail archive of the list.

Unlike the previous case analysis for the Conceptual Graph community (Gaines, 1993b; Gaines, Chen & Shaw, 1997) in Chapter 7, the identities of members in the Gbrain community will be kept confidential. The reason for anonymity in this case is due to possible ethical issues involving social-psychological rating in the SYMLOG methodology. Although the Gbrain list is in the public domain as is the Conceptual Graph list, the group-dynamic analysis in the SYMLOG methodology requires rating judgment. It is best not to risk offending the judged parties. Since the purpose here is to demonstrate ListA and the methodologies, anonymity would not effect the patterns of social network structure and group dynamics, which are our main interests.

## 8.4.1 Evolution of the Global Brain Special Interest Community

| Subject Line | Postings |
|---|---|
| Re: global brain conference | 6 |
| Re: The Life of Brain | 6 |
| Re: Super organisms - sane or insane | 4 |
| Re: Opening up the Gbrain list | 4 |
| Re: Dreaming of Reality | 4 |
| Re:  Wired parodies the evolution of the super-organism ;-) | 3 |
| cybernetic immortality | 3 |
| Re: [Fwd: howdy fellow global-brainers -- are you still out | 3 |
| Re: super-systems, super-systems & co | 2 |
| Re: self and mind | 2 |
| Re: Semantic web vision [from Sasha Chislenko] | 2 |
| Re: John Earls on a Super-brain model of Andean communities | 2 |
| Re: Evaluation submissions to join Gbrain-l | 2 |
| Re: Opening up the Gbrain list [J. de Rosnay] | 2 |
| Re:  Super-organisms - sane or insane | 2 |
| Organization of the Global Brain group | 2 |
| Re: New subscriber: Andy Edmonds | 2 |

**Table 14 Posting Frequency per Topical Thread [postings > 1]**

After converting the Gbrain list into HTML formats by the Hypermail program, we can obtain useful data files by running the PREP tool in ListA. One useful statistics table is the *posting frequency per topical thread* (Table 14) derived from the *Subject ID* data file. From this table, we can see which topical threads have captured the interests of the community. In this case, the global brain conference and the life of brain seem to capture the members' imagination.

**Total No. of Days = 316; No. of Messages = 111; Days with Postings = 50**

…

| DAY # | DATE | Number of Postings | As Percentage of Total Postings |
|---|---|---|---|
| **D193** | 97/3/7 | 1 | 0.9% |
| **D214** | 97/3/28 | 1 | 0.9% |
| **D273** | 97/5/26 | 1 | 0.9% |
| **D274** | 97/5/27 | 1 | 0.9% |
| **D275** | 97/5/28 | 4 | 3.6% |
| **D276** | 97/5/29 | 6 | 5.4% |

…

**Table 15 Partial List of Message Frequency Table**

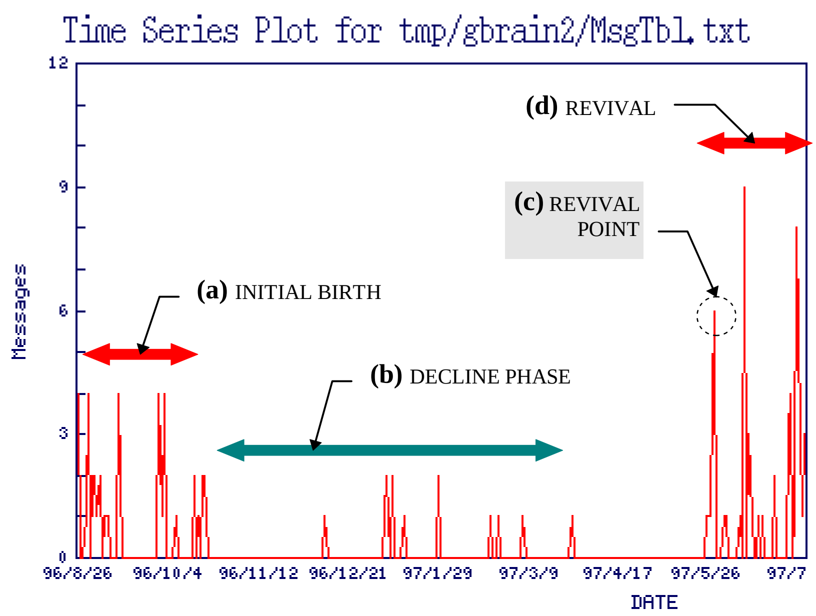


**Figure 26 Time-series Plot for the Global Brain Community**

After the initial data preparation by PREP, we can gauge the life cycle of the Gbrain community and its vitality by running the TIME SERIES tool. The resulting time-series plot (Figure 26) tells us the life story of the community. We see that an initial active posting pattern in its (a) *Initial Birth* phase, follows by relative quietness in the (b) *Decline Phase*, then suddenly around a critical (c) *Revival Point*, the list has been resuscitated into a (d) *Revival*.

The revival point in the time-series plot prompts an interesting venue for further investigation. Why did it occur at that specific period? By studying the Message Frequency Table (Table 15), we can see that in the two months following 97/3/28 there was no activity in the list. The list was almost starved to death.

We can also compute the autocorrelation statistics (e.g., ACF lag-1 function) for the list. There are N=49 pairs of intervals with the total sum of 111 messages, thus $\overline{X}$ = 111/49 = 2.265. From Equation 3, we derive an ACF lag-1 value of −0.0168. The standard error (computed from Equation 4) is SE lag-1 ACF = 0.1429. Hence, the critical values for the rejection of the null hypothesis of zero autocorrelation are ±2(SE) = ±0.2858. Since the observed ACF is −0.0168, which is neither greater than +0.2858, nor smaller than −0.2858, it is not significantly different from a zero ACF. Therefore in this case, we fail to reject the null ACF hypothesis.

As stated earlier in Chapter 1, there is a range of stability for each of the numerous variables in all living systems. Ordinarily, there is a standard range of rates at which input enters a system. If the input rate falls below this range, it constitutes a *lack stress*. One class of stress is the *information stresses*, including: information input lack or underload, resulting from a dearth of information in the environment or from improper function of the external sense organs or input transducers (Miller, 1978).

Adaptation to change, when the basic aim is to maintain the health and integrity of the system, falls within the maintenance imperative. For instance, the maintenance of the steady state of message flow is essential for the health and vitality of a listserver-based virtual community. A virtual community needs to monitor and regulate the information flow among its members. It has to adapt to the changing patterns of topical interests and social behaviours of its members. Message overflow or underflow on a list can induce information stresses and strains to the well-being of the system.

After a careful re-examination of the Hypermail list archive at the (c) Revival Point, the list did, indeed, go through a fundamental change. Before that point, the Gbrain list had around 10 members, and the community had a tight *boundary* subsystem in the sense that one had to be invited to become a member. On 97/5/23, in a topical thread: ‘Opening up the Gbrain list’, an influential member of the list (User ID: U2) suggested that the community should change its policy for admitting members; and its follow up responses were quite favourable. The consensus was to adopt a *peer-review* system (Harnad, 1994; Harnad, 1995) for new applicants.

Consequently, after the critical revival point (D273-D276), the list has come alive again through its adaptation to change. Gradually, the list reached its current membership of 17 participants.

8.4.2 Social Networks in the Global Brain Community

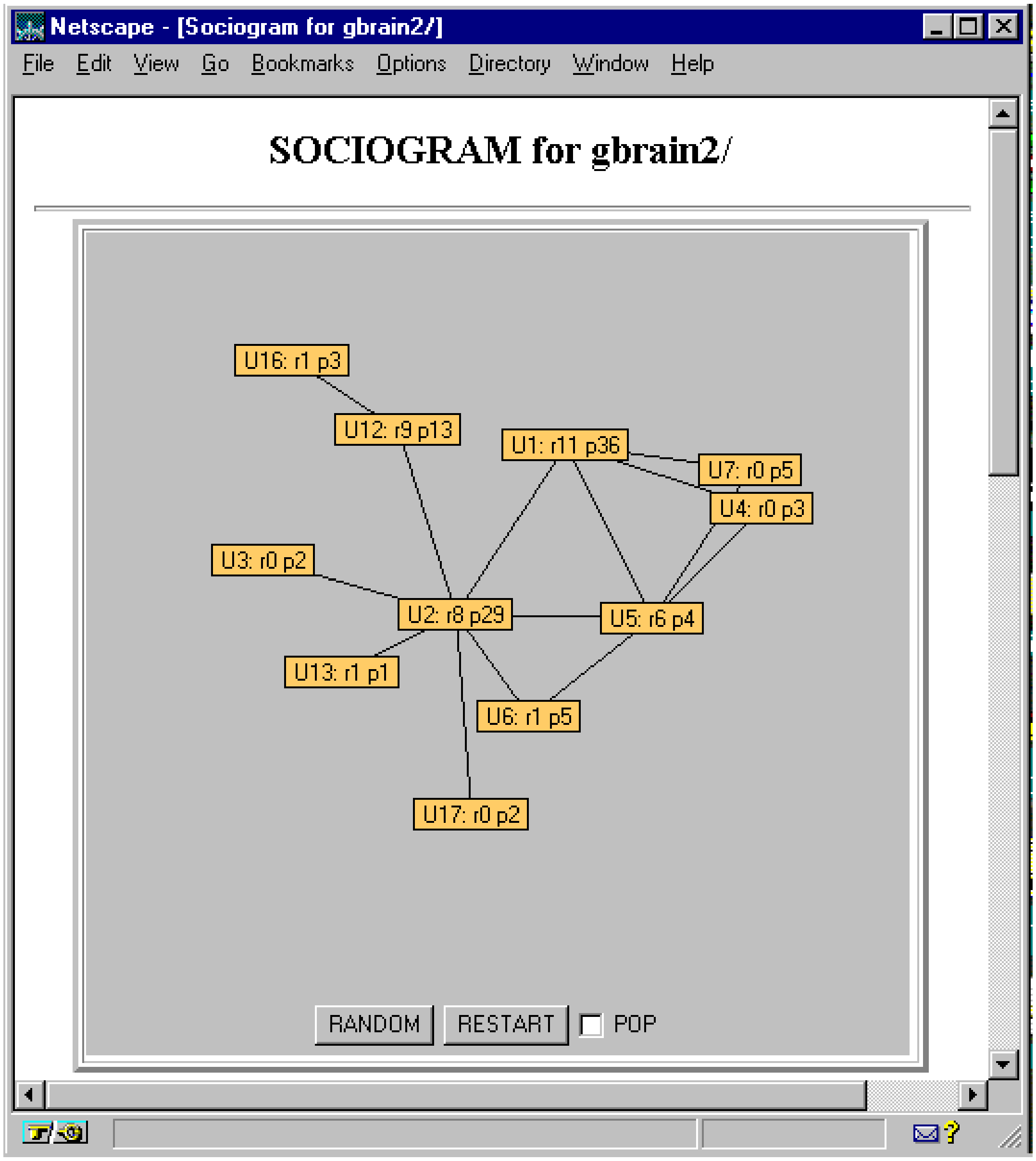


**Figure 27 Dynamic Sociogram of the Global Brain Community**

We can run the SOCIOGRAM tool in the ListA to get a first impression of the Gbrain community. The resulting dynamic interactive sociogram (Figure 27) shows its active social network. From the sociogram we can see that 3 members (U1, U2, U5) have formed tight bonds among them. They clearly are the influential members in the Gbrain social network, since they have higher numbers of arcs linked to them than others.

**tmp/gbrain2/Matrix.txt**

TOTAL MSGS N=111 Posters=17

| U_ID | 1 | 12 | 2 | 5 | 16 | 6 | 13 | 10 | 17 | 14 | 7 | 11 | 15 | 9 | 3 | 4 | 8 | TOTAL | POSTS |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| U 1: | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | U 1 | 36 |
| U 2: | 7 | 4 | 3 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | U 2 | 29 |
| U 12: | 0 | 5 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | U 12 | 13 |
| U 7: | 1 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | U 7 | 5 |
| U 6: | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | U 6 | 5 |
| U 5: | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | U 5 | 4 |
| U 16: | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | U 16 | 3 |
| U 14: | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | U 14 | 3 |
| U 4: | 2 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | U 4 | 3 |
| U 17: | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | U 17 | 2 |
| U 3: | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | U 3 | 2 |
| U 11: | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | U 11 | 1 |
| U 10: | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | U 10 | 1 |
| U 13: | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | U 13 | 1 |
| U 15: | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | U 15 | 1 |
| U 9: | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | U 9 | 1 |
| U 8: | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | U 8 | 1 |
| U_ID | 1 | 12 | 2 | 5 | 16 | 6 | 13 | 10 | 17 | 14 | 7 | 11 | 15 | 9 | 3 | 4 | 8 | | |
| Cited: | 11 | 9 | 8 | 6 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | |

**Figure 28 Sorted Social Matrix of the Global Brain Community**

By examining the sorted Sociomatrix date file (Figure 28) and the User ID data file (partially listed in Table 16), we can study who have contributed the most, in absolute, terms to the Gbrain community.

| User ID # | Postings | Cited by Others |
|---|---|---|
| **1** | 36 | 11 |
| **2** | 29 | 8 |
| **12** | 13 | 9 |
| **7** | 5 | 0 |
| **6** | 5 | 1 |
| **5** | 4 | 6 |
| **16** | 3 | 1 |
| **14** | 3 | 0 |
| **4** | 3 | 0 |
| **17** | 2 | 0 |
| **3** | 2 | 0 |
| **11** | 1 | 0 |
| **10** | 1 | 0 |
| **13** | 1 | 1 |
| **15** | 1 | 0 |
| **9** | 1 | 0 |
| **8** | 1 | 0 |

**Table 16 Posting and Cited Frequencies in the Global Brain List (Sorted)**

An interesting observation is that the members with more contributions to the list also received higher citing frequencies. Such a positive correlation could be, in effect, evidence for positive feedback in the community.

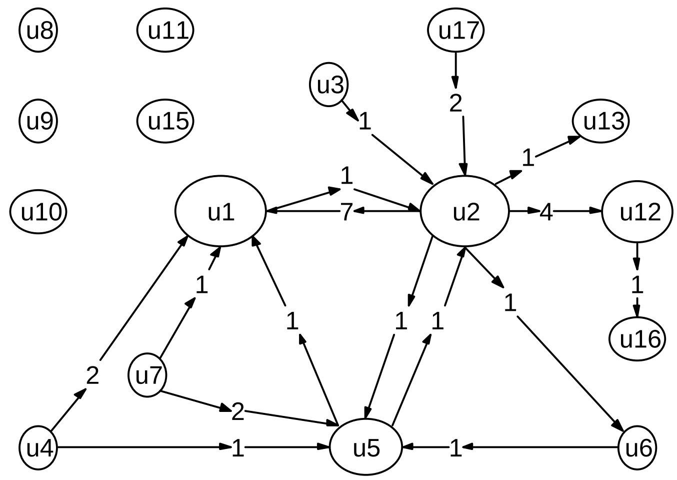


**Figure 29 Gbrain Sociogram Based on its Sociomatrix**

| User ID # $n_i$ | Outdegree $D_O(n_i)$ | Indegree $D_I(n_i)$ | Degree Centrality $C'_D(n_i)$ | Degree Prestige $P'_D(n_i)$ |
|---|---|---|---|---|
| **1** | 1 | 4 | .0625 | .2500 |
| **2** | 6 | 5 | .3750 | .3125 |
| **3** | 1 | 0 | .1660 | 0 |
| **4** | 2 | 0 | .1250 | 0 |
| **5** | 3 | 5 | .1875 | .3125 |
| **6** | 1 | 1 | .0625 | .1660 |
| **7** | 2 | 0 | .1250 | 0 |
| **8** | 0 | 0 | 0 | 0 |
| **9** | 0 | 0 | 0 | 0 |
| **10** | 0 | 0 | 0 | 0 |
| **11** | 0 | 0 | 0 | 0 |
| **12** | 2 | 2 | .1250 | .1250 |
| **13** | 0 | 1 | 0 | .0625 |
| **14** | 0 | 0 | 0 | 0 |
| **15** | 0 | 0 | 0 | 0 |
| **16** | 0 | 1 | 0 | .0625 |
| **17** | 1 | 0 | .0625 | 0 |

**Table 17 Degree Indices**

From the Gbrain sociomatrix (Figure 28), we can draw a *detailed sociogram*[6] (Figure 29) and compute useful indices, such as: outdegree (Equation 7), indegree (Equation 1), degree centrality (Equation 10) and degree prestige (Equation 11). Those computed degree indices are listed in Table 17. From Equation 9, we can calculate the density of the social network, the Gbrain has a density of $\Delta = 0.0699$.

Based on these indices, we can see that member U2 has the highest degree centrality ($C'_D(n_2) = .3750$). Two members, U2 and U5, have the highest degree prestige ($P'_D(n_2) = P'_D(n_5) = .3125$). Together with U1, who has the second highest degree prestige ($P'_D(n_1) = .2500$), they have the highest degree of influence over others. After careful analysis of their postings, these three members seem to have formulated an *invisible college* (Crane, 1972). This subgroup in the Global Brain community has often brought in postings or/and research findings from other listservers of related disciplines to the community. As stated in Chapter 2, under the leadership of those members, the subgroups of collaborators recruit and socialize new members and try to maintain a sense of commitment to the research area among members, thus the formation of *solidarity groups* (Mullins 1968).

---

[6] Concept mapping tools (Gaines & Shaw, 1995; Kremer, 1997; Flores-Mendez, 1997) are ideal for assisting sociogram construction.

## 8.4.3 SYMLOG Dimensions of the Global Brain Community

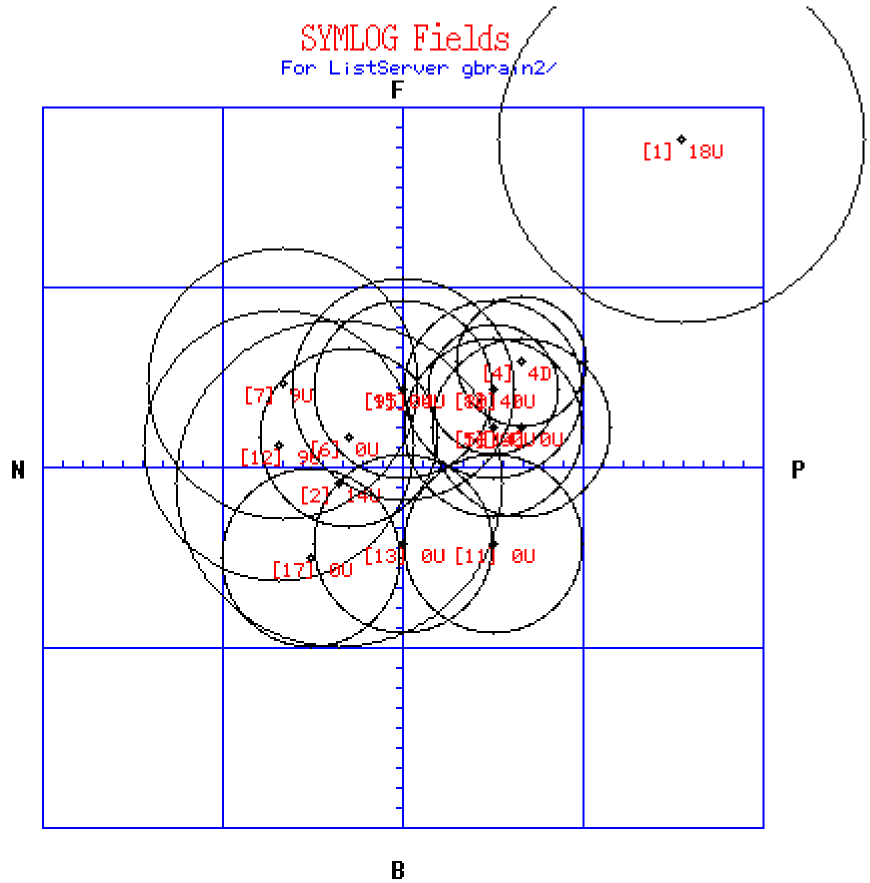


**Figure 30 Original Field Diagram of the Global Brain Community**

After we have encoded SYMLOG interaction scores using the ENCODE SYMLOG tool in ListA, we an proceed to analyze the social-psychological dynamics of the Gbrain community.

When ListA makes a Group Average Field Diagram after we have pressed the FIELD DIAGRAM, many differences in the way particular members ‘act’ cancel each other, and the average location of the image points tend to regress toward the zero point of the Diagram. Group Average Field Diagrams, in particular, need to be expanded or contrasted before they can be compared optimally to the relative distances among member's particular Fields.

It is possible that some plotting scores from the formula will not be contained within the Field Diagram scale limit of ±18 as in the case member of U1 in Figure 30. If the absolute scores go higher than 18, the application of the Expansion Multiplier to the original Field Diagram is necessary.

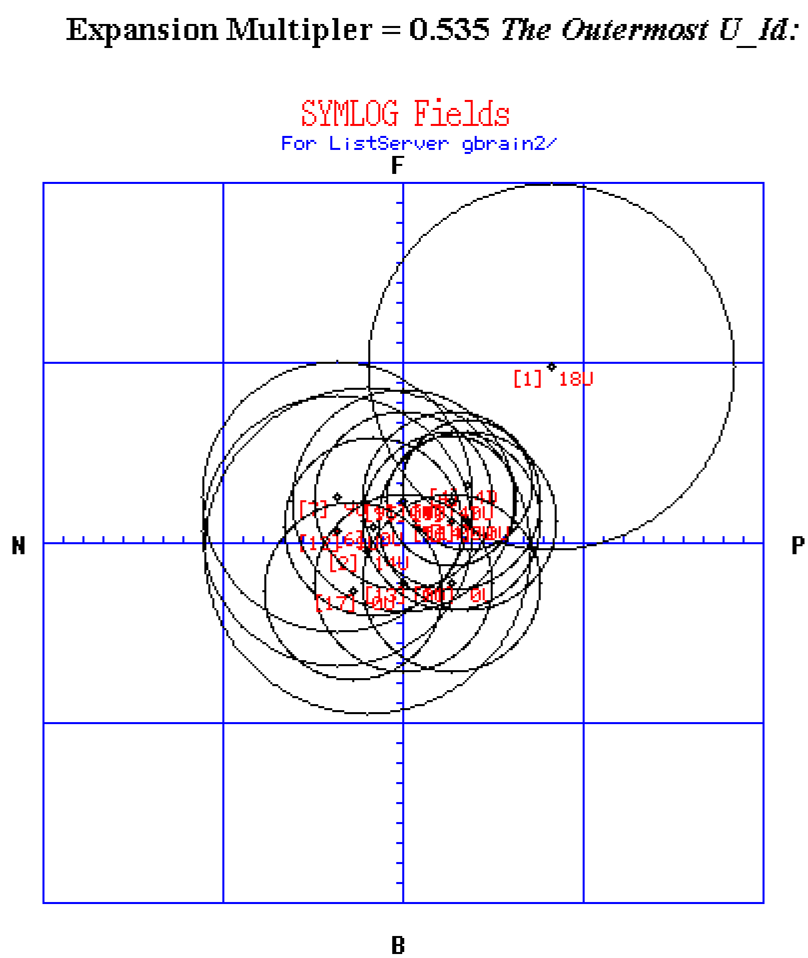


**Figure 31 Adjusted Original Field Diagram of the Global Brain Community**

After applying the Expansion Multiplier Factor of 0.535 to Figure 30, we get a resulting Adjusted Field Diagram depicted in Figure 31. Member U1 is clearly an *outlier* in the SYMLOG data set. An outlier is an extreme measurement that stands out from the rest of the sample. From the adjusted original field diagram and SYMLOG Interaction Scores (Table 18), we see that U1 is the most dominant figure in the Gbrain community (UD = 18.89). Coincidentally, U1 is the founder of the Gbrain list.

| | | | |
|---|---|---|---|
| Summary[1]<br>B:7 F:29 N:3 P:12 D:1 U:12 | **FB: 16.3773016361116** | PN: 13.9817321414819 | **UD: 18.8989595432356** |
| summary[2]<br>B:14 F:13 N:10 P:8 D:3 U:12 | FB: -0.84945271939537 | PN: -3.19969558771842 | **UD: 14.8616559977764** |
| summary[3]<br>B:0 F:1 N:0 P:1 D:1 U:0 | FB: 3.88303853556602 | PN: 4.59610177413889 | UD: -4.7004369611246 |
| summary[4]<br>B:0 F:3 N:0 P:2 D:1 U:0 | FB: 5.25575017197957 | PN: 5.99552083091854 | UD: -4.7004369611246 |
| summary[5]<br>B:1 F:2 N:0 P:1 D:1 U:1 | FB: 1.97142540984253 | PN: 4.59610177413889 | UD: 0 |
| summary[6]<br>B:2 F:3 N:2 P:1 D:1 U:1 | FB: 1.49246679292548 | PN: -2.63578285721965 | UD: 0 |
| summary[7]<br>B:1 F:4 N:2 P:0 D:0 U:4 | FB: 4.10904562676115 | PN: -5.99552083091854 | UD: 9.21169969242068 |
| summary[8]<br>B:0 F:1 N:0 P:1 D:1 U:0 | FB: 3.88303853556602 | PN: 4.59610177413889 | UD: -4.7004369611246 |
| summary[9]<br>B:0 F:1 N:0 P:0 D:0 U:0 | FB: 3.88303853556602 | PN: 0 | UD: 0 |
| summary[10]<br>B:0 F:1 N:0 P:1 D:0 U:0 | FB: 3.88303853556602 | PN: 4.59610177413889 | UD: 0 |
| summary[11]<br>B:1 F:0 N:0 P:1 D:0 U:0 | FB: -3.88303853556602 | PN: 4.59610177413889 | UD: 0 |
| summary[12]<br>B:6 F:7 N:4 P:1 D:0 U:4 | FB: 1.0132522017913 | PN: -6.11361497929361 | UD: 9.21169969242068 |
| summary[13]<br>B:1 F:0 N:0 P:0 D:0 U:0 | FB: -3.88303853556602 | PN: 0 | UD: 0 |
| summary[14]<br>B:1 F:2 N:0 P:2 D:0 U:0 | FB: 1.97142540984253 | PN: 5.99552083091854 | UD: 0 |
| summary[15]<br>B:0 F:1 N:0 P:0 D:0 U:1 | FB: 3.88303853556602 | PN: 0 | UD: 4.7004369611246 |
| summary[16]<br>B:1 F:2 N:0 P:1 D:0 U:0 | FB: 1.97142540984253 | PN: 4.59610177413889 | UD: 0 |
| summary[17]<br>B:2 F:0 N:1 P:0 D:0 U:0 | FB: -4.5693943537728 | PN: -4.59610177413889 | UD: 0 |

**Table 18 SYMLOG Interaction Scores**

By examining the total number of messages posted by U1 (36/111= 32.4% of the list) and his **Forward-Backward** dimension (FB[1] = 16.38 from Table 18 and Figure 31), we can see that U1 (the founder of the list) is indeed the *de facto* task leader in the community. Both $T_L(n_1) = 36$ and FD[1] = 16.38 are the highest among U1's peers. In addition, his degree prestige $D'_P(n_1) = .2500$ is the second highest ranking. After re-examining the contents of his messages, U1 has contributed many task-oriented postings to the list indeed.

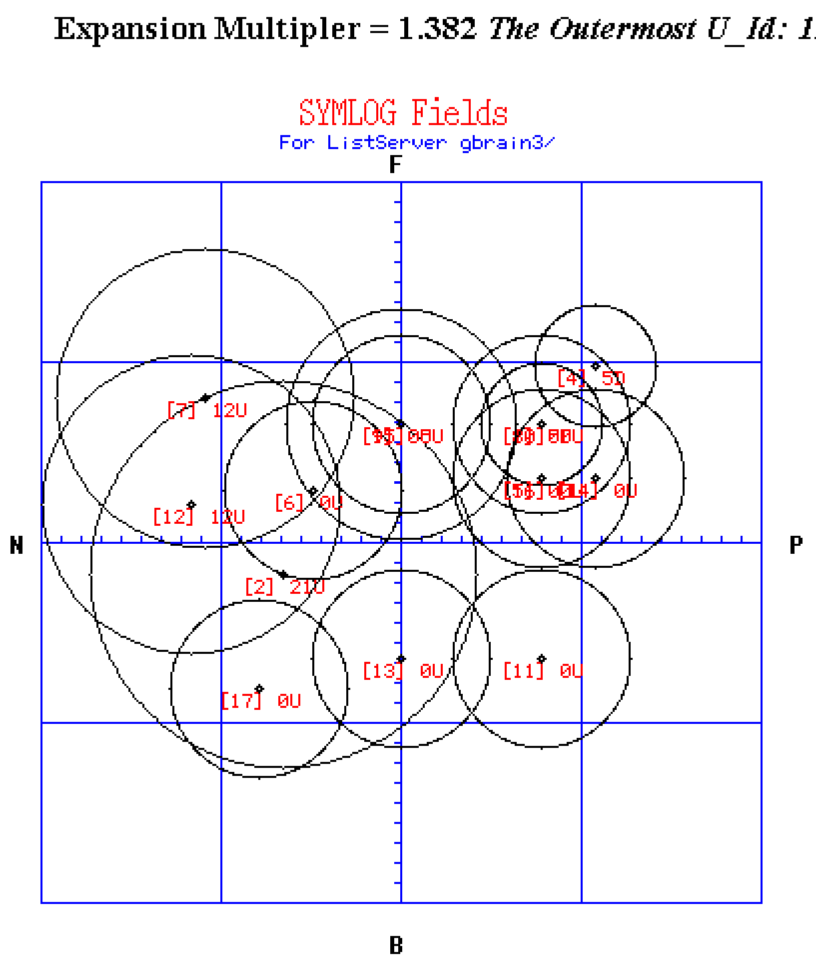


**Figure 32 Adjusted Gbrain Field Diagram (with Outlier Removal)**

According to standard statistics practice, removal of an outlier from data set often reveals more information about other data points (McClave & Dietrich, 1991). Figure 32 depicts the Adjusted Field Diagram after U1 has been removed from the data set. We can see that member U2 is clearly the most dominant member right after U1 (UD[2] = 14.86). From the diagram, U2 seems to the central point of among more active members (U7, U12, U6, U17, U5). From analyses of his messages, U2 tends to act as the *de facto* socio-emotional leader in the community. His messages often contained emotional lively and charged discourses. Since his FB[2]: -0.849 is not the lowest (U17 has the lowest Backward score: FB[17] = -4.56), we know he is not the most emotional member. Nevertheless, U2 is still among the only four (U2, U17, U13, U11) out 17 members who exhibit socio-emotional direction (i.e., with negative FB score). In addition, from the social network analysis, he has the highest degree centrality ($C'_D(n_2) = .3750$) and degree prestige ($P'_D(n_2) = .3125$). Therefore, we can conclude that members U2 and U1 are the socio-emotional and the task-oriented leaders respectively.

## *8.5 Summary*

The importance of coding and analysis of group-dynamics in order to better understand CSCW has long been recognized. It has been also acknowledged that using such analysis for feedback purposes can enhance cooperative behaviours (Losada & Markovitch, 1990). Automated analyzers, likes ListA, allow us to receive faster feedback from analyses of virtual cooperative interaction in special interest communities.

This chapter introduces ListA: a Listserver Analysis program and a sample case special interest community—the Global Brain Community—for demonstrations of ListA and the systematic methodologies described in Chapter 7.

ListA consists of 5 analysis tools: (1) PREP: for preparation of data files; (2) TIME SERIES: for time-series plot; (3) SOCIOGRAM: for generation of dynamic interactive sociogram; (4) ENCODE SYMLOG: for rating SYMLOG interaction scores; (5) FIELD DIAGRAM: for group average field diagrams (original & adjusted)

First, we run ListA to prepare data files for later analyses. ListA then generates a time-series plot for us to examine a list's life cycle, and a dynamic interactive sociogram enabling us to investigate its social networks. In addition, ListA offers input-facilitation of SYMLOG coding schemes in respect to individual postings. After the coding phase has been completed, ListA allows researchers interactively to view SYMLOG field diagrams for further analysis.

In the last section of the chapter, the Global Brain special interest community offers as a test case for analyses of its life cycle, evolution, social structure, member contribution, centrality, influence, leadership and group dynamics.

# CHAPTER 9

# Tracking Reproduction, Migration and Dissemination

In the last two chapters, we have seen methodologies and techniques for investigating the social-technical dynamics at the level of special interest communities. In this chapter, our focus will be on techniques and methodologies for tracking vehicle/memetic reproduction, migration and dissemination. Here, the focal level of the investigation will be the Internet community at large.

"How can we track diffusion processes of software or innovation on the Internet?" is a common question concerning market researchers and social scientists alike (Hoffman & Novak, 1995). Diffusion is the process by which an innovation is communicated through certain channels, over time, among the members of a social system. It is a special type of communication, in that the messages are concerned with new ideas, i.e., novel memes (Rogers, 1995).

In the following sections, I will discuss techniques and methodologies for tracking diffusion processes of software and webpages on the web/net. CHRONO will be used as a sample case to illustrate those techniques and methodologies. The fundamental philosophy behind the methodologies in this chapter is that they attempt to use existing, public-domain constructs and tools as much as possible. Furthermore, these methodologies should be ethical and not to infringe individual privacy. Thus, they should be as least intrusive as possible.

## *9.1 Tracking Techniques and Methodologies*

As stated by Miller (1978), the third imperative of living systems is *propagation of the system through reproduction and/or dissemination*. Each gene strives to perpetuate its peculiar pattern. The current survival vehicle cannot be maintained forever, but in theory the gene or meme can. Reproduction is the ultimate mechanism for genetic survival. For memes, dissemination serves a similar function. The more widespread an idea is, the more likely it will survive.

Adapting a gene/meme centric perspective (Williams, 1966; Dawkins, 1989b), software such as CHRONO is analogous to a survival vehicle for a community of memes unified by its dominant meme for *chronological awareness maintenance* (other co-memes are: *server-side responsibility* and *browsing strategy*). Organisms struggle through life in order to survive and to pass on their genes through reproduction. Whenever there are opportunities, a species can extend its survival chance through migration to new territories (Wilson, 1992). Similarly CHRONO can be considered as a particular species of chronological awareness maintenance mechanisms. Its purpose, as stated in the third imperative of life, is to propagation the system through reproduction and/or dissemination.

CHRONO reproduces like a flower spore and migrates to different websites in cyberspace. Listserver/newsgroups, category indexes, and search engines (e.g., www-talk, Yahoo, Alta Vista) are like memetic pollinators in their ability to disseminate the '*chronological awareness mechanism*' meme. Anyone of those services can inject the meme into the mind of a susceptible human agent (for example a system administrator); she then downloads a copy of CHRONO to her website and configures the system; thus, the process gives birth to another CHRONO mechanism migrated to her website (i.e., a new territory). Once it is established in a new location, another system administrator who visits the website may discover CHRONO and finds it useful, hence the life cycle continues.

CHRONO thus offers an ideal case for conducting an observational experiment for tracking software diffusion processes on the net. In addition to introducing tracking techniques and methodologies, the following subsections also describe an observational experiment about the diffusion patterns of CHRONO. The experiment commenced on

May 13, 1996; and continued till April 30,1997. It was designed as an example case to demonstrate systematic methodologies for investigation of reproduction, migration and dissemination processes in the cyberorganism.

The following sections introduce techniques and methodologies such as: homing beacon; genealogy tracer; and meta-monitor. First, we need a technique for tracking software activities.

### *9.2 Homing Beacon*

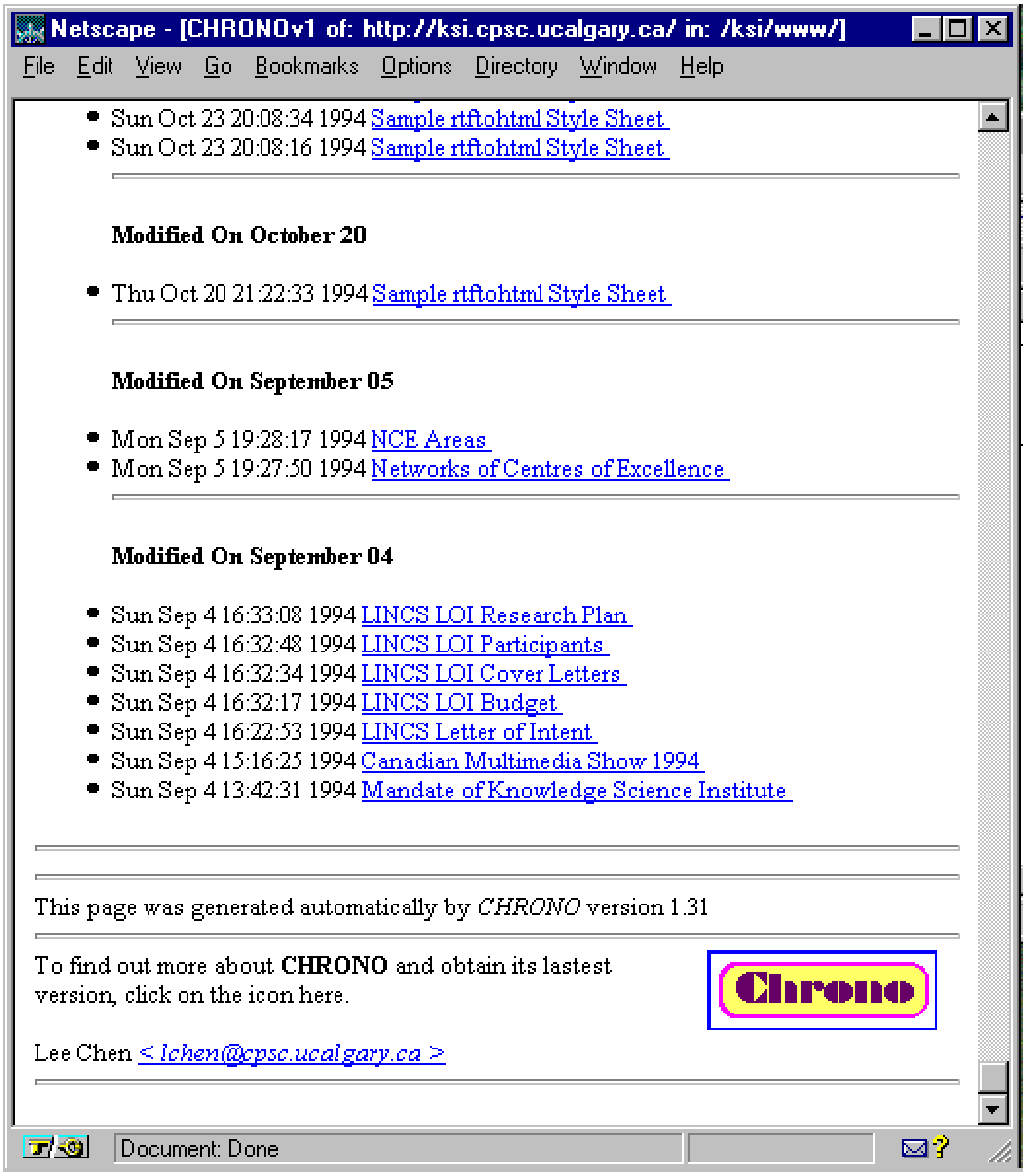


**Figure 33 Bottom of a CHRONO Generated HTML Index**

Ideally, we should try to use basic constructs available to us on HTTP servers for tracking software activities on the web. In ecological field studies involving long-distance tracking of migrating birds, ecologists sometimes utilize radio *homing beacons*, attached to migrating birds, to tell us where the birds are and how they are doing (Gotelli, 1995).

Those homing devices should be non-intrusive to the studied subjects as much as possible.

We can use a simple technique for achieving a similar effect of tracking software. Embedded at the bottom right hand corner of every chronological awareness page is a **CHRONO** GIF Icon (Figure 33). Below is its corresponding HTML code (Figure 34) generated by all issued copies of CHRONO, except in-house copies residing at the Knowledge Science Institute (KSI) HTTP server:

```
…
<HR><HR>This page was generated automatically by
<I>CHRONO</I> version 1.31<HR>
<A href="http://ksi.cpsc.ucalgary.ca:8008/cgi-bin/release?e1">
<img src=
     "http://ksi.cpsc.ucalgary.ca:8008//CHRONO/gif/chrono1.gif"
     align=right hspace=10 alt="CHRONO ICON">
</A>
To find out more about <b>CHRONO</b> and
obtain its lastest version, click on
the icon here. <BR>
…
```

**Figure 34 HTML Code Fragment of CHRONO page**

When someone checks out a *What's New* page generated by CHRONO on a website somewhere, its web server fetches the specified CHRONO GIF image from the KSI web server. Such a fetching action generates an 'HTTP hit' and record it in KSI's HTTP *access_log* and *referer_log* (both are standard log files available on the NCSA's web server).

```
sunsite.wits.ac.za - - [25/Jun/1996:00:42:57 -0600] "GET /CHRONO/gif/chrono1.gif HTTP/1.0"
304 0
```

**Figure 35 HTTP access_log**

```
http://sunsite.wits.ac.za/htdocs_chrono.html -> /CHRONO/gif/chrono1.gif
```

**Figure 36 HTTP referer_log**

For example, two entries in the access_log (Figure 35) and referer_log (Figure 36) tell us that a visitor of the *sunsite.wits.ac.za* website checked out a CHRONO generated *What's New* page named *htdocs_chrono.html* on June 25, 1996. This information is useful in compiling usage statistics that tell us how well the CHRONO species as a whole is doing. In addition, the information gathered does not violate any user privacy, since it contains no knowledge of user identity.

In short, the technique of embedding a tracking GIF icon in software acts like placing a homing beacon for software usage tracking methodology. The GIF should be strategically placed (like at the bottom of the page), so that users do not need to wait for its fetching process and be able to read the page without inconvenience.

## 9.2.1 CHRONO Usage

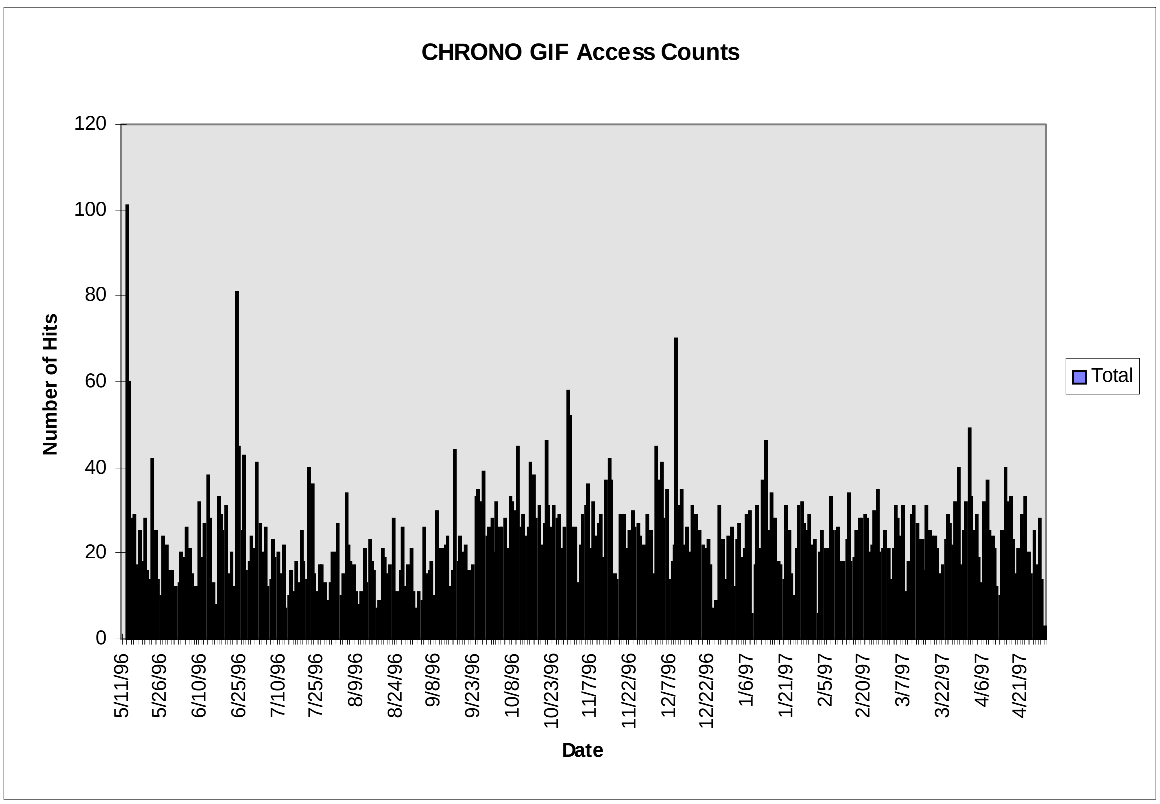


**Figure 37 CHRONO Usage**

We can use a tracking web site's access_log (e.g., KSI's) to compile a daily usage graph for the tracked software. By using a homing beacon GIF, we can obtain activity counts in order to determine the well being of a particular software species. Figure 37 shows daily usage patterns of CHRONO on the Internet community at large, from May 13, 1996 through April 30, 1997. From the referer_log, we can count the number of unique website pages that have activated the homing beacon. In total, there are at least 137 CHRONO pages available on the web (the figure was complied on April 30, 1997).

For our experimental purpose, there were only three public announcements made about CHRONO (96/5/13, 96/08/01, and 96/09/09). They were designed to test the effectiveness of memetic dissemination channels[7]. From the graph, we see that the highest hit rate occurred around the first day of CHRONO public release (well above 100 hits). It fits well with the *novelty effect*—that is, a new idea or novel meme often has the highest impact when people encounter it the very first time.

[7] The next section introduces techniques for tracing which channels are more effective.

## *9.3 Genealogy Tracer*

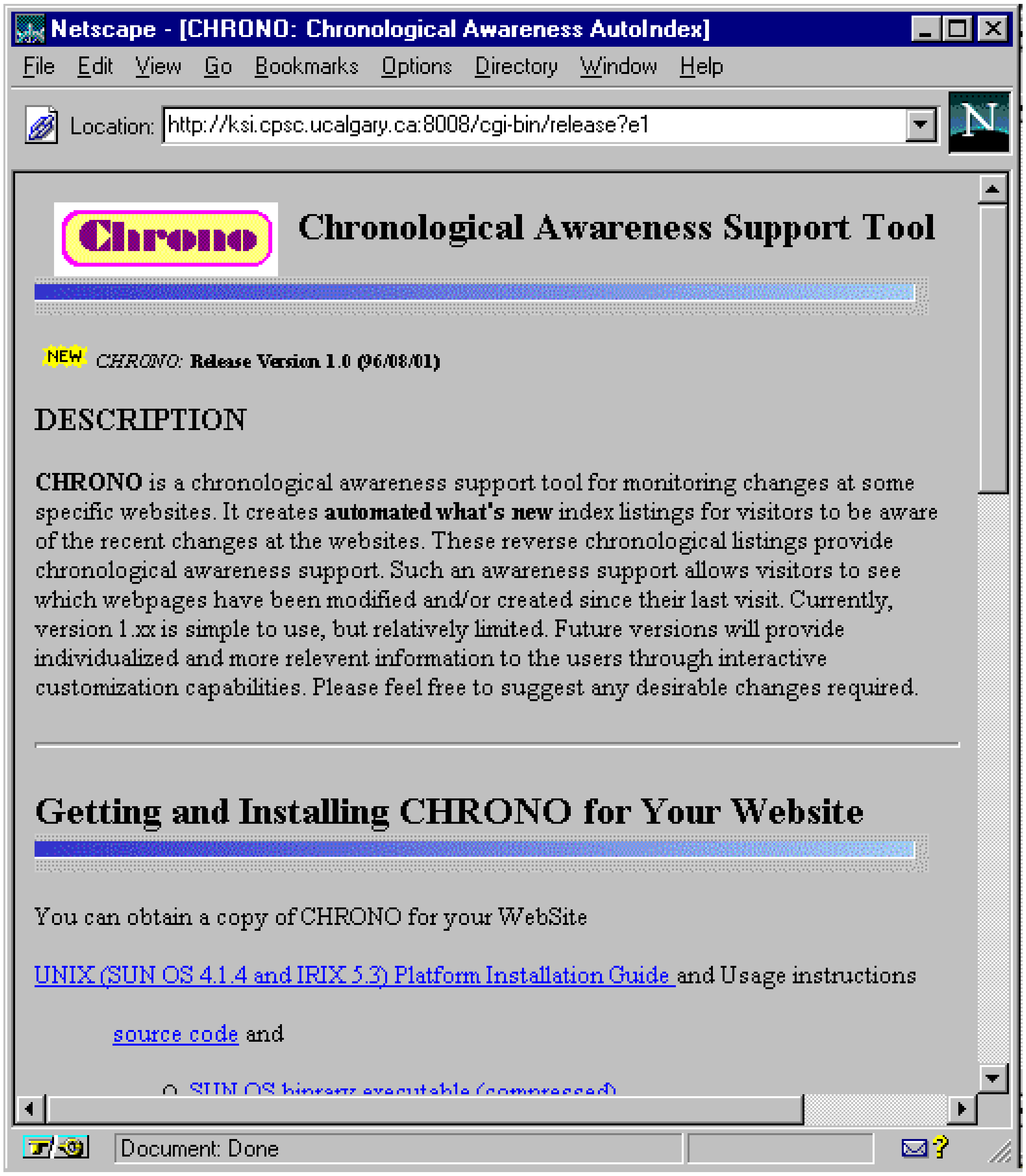


**Figure 38 CHRONO Release Page**

In addition, to being able to track an organism's whereabouts, it is a good idea to be able to trace the genealogy of organisms as a species as they propagate and migrate to other regions. Similarly, we would like to identify what is the lineage of a particular copy of

CHRONO. More precisely we would like find out where, and how, someone discovered CHRONO.

Figure 38 is a screen shot of CHRONO Release Page, this is where interested parties go to find instructions for downloading and setting up a copy of CHRONO. This page is generated through the *genealogy tracer* script in which it places a unique identifier in the HTTP referer_log.

By clicking at the CHRONO Icon (a hyperlink) as depict in Figure 33, an interested user invokes a genealogy tracer CGI Perl script located at KSI which returns a customized Release Page. In this case, the hyperlink specified in the HTML (Figure 34):

```
<A href="http://ksi.cpsc.ucalgary.ca:8008/cgi-bin/release?e1">
```

returns a Release Page embedded the "e1" tracer code. Associated with this customized Release Page are hyperlinks to CHRONO binary and source code which also contain the same tracer code "`e1`".

The genealogy tracer technique can be used in our systematic methodology for tracking memetic transmission. For example, we can determine where someone had heard or learned about CHRONO. Figure 39 depicts partial sections of a Software Release Table with tracer codes. I have used these codes to track whether visitors discovered CHRONO through the www-talk listserver, Yahoo, or other channels.

```
1996 05 13 www-talk    To: www-talk@w3.org
  http://ksi.cpsc.ucalgary.ca:8008/cgi-bin/release?tw1
  (page to http://ksi.cpsc.ucalgary.ca:8008/cgi-bin/release?e1)
1996 05 16 yahoo (1st try, not indexed):
 Computers and Internet/Internet/World Wide Web/Searching the Web/Search Engines
 http://ksi.cpsc.ucalgary.ca:8008/cgi-bin/release?1yh
1996 06 27 yahoo: (2nd try, still not indexed)
 Computers and Internet/Internet/World Wide Web/Searching the Web/Search Engines
 http://ksi.cpsc.ucalgary.ca:8008/cgi-bin/release?2yh
 "An Automated "WHAT's NEW" Chronological Awareness Engine."
1996 08 01 www-talk To: www-talk@w3.org
  http://ksi.cpsc.ucalgary.ca:8008/cgi-bin/release?tw2
  (with 2nd general release. /CHRONO/gif/chrono2.gif)
  (page to http://ksi.cpsc.ucalgary.ca:8008/cgi-bin/release?e2)
  [access_log 1996 08 01 5:28pm]
1996 09 09 yahoo: (3nd try, worked this time)
 Computers and Internet/Internet/World Wide Web/Searching the Web/Search Engines
 http://ksi.cpsc.ucalgary.ca:8008/cgi-bin/release?3yh
 "CHRONO Automated "WHAT'S NEW" Chronological Awareness Engine."
```

**Figure 39 Tracer Codes for Release Lineages**

By analyzing the HTTP access_log, we can discover when someone accesses the Release Page, downloads a copy of CHRONO (either binary or C source), and more importantly, where she learned about CHRONO. For example, Figure 40 is a Release Page entry in KSI access_log:

```
sunsite.nstu.nsk.su - - [25/Jun/1996:04:32:25 -0600] “GET /cgi-bin/release?tw1 HTTP/1.0”
200 4071
```

**Figure 40 Genealogy Tracer of the First WWW-Talk Announcement**

The entry tells us that someone from the sunsite.nstu.nsk.su site accessed the Release Page on June 25, 1996. With a high degree of certainty, this person probably had learned about CHRONO from the first announcement of CHRONO (May 13, 1995) on the www-talk[8] listserver.

The genealogy tracer technique can be used in conjunction with the homing beacon technique to trace migration patterns. For instance, we can use a different CHRONO GIF for each genealogy. In our CHRONO migration study, there are two types of homing beacon frequencies:

```
“http://ksi.cpsc.ucalgary.ca:8008//CHRONO/gif/chrono1.gif”
“http://ksi.cpsc.ucalgary.ca:8008//CHRONO/gif/chrono2.gif”
```

The first one tracks the www-talk lineage (release codes: `tw1` and `tw2`) and the second one tracks the Yahoo lineage (release code: `3yh`).

The genealogy tracer technique of using a CGI Release Page for tagging, together with the systematic strategy of selective announcement provide us a methodology to investigate the effectiveness of different diffusion channels and to analyze the resulting pattern of memetic awareness in the cyberorganism.

## 9.3.1 CHRONO Reproduction and Migration Patterns

The CHRONO diffusion tracking experiment includes three phases:

---

[8] WWW-Talk can be reached via: www-talk@w3.org

1. 96/05/13: 1st WWW-Talk announcement (SUN OS release)
2. 96/08/01: 2nd WWW-Talk announcement (SGI IRIX release)
3. 96/09/09: 3rd Yahoo Registration (successful attempt)

The first CHRONO release public announcement was made to the www-talk mailing list, an active list dedicated to discussion about the web. The release included the genealogy code `tw1`. A second announcement in www-talk offered a SGI IRIX binary version of CHRONO instead of SUN OS binary, otherwise it was exactly the same as the first announcement. The third release was made through Yahoo[9]. It is a web indexer service, a different dissemination channel than listservers like www-talk. There were also two attempts to register CHRONO on the Yahoo website in May and June, unfortunately operators at Yahoo Inc. did not follow up both times.

Figure 41 depicts the diffusion pattern of CHRONO through release and download processes.

[9] Yahoo search indexer, via http://yahoo.com/

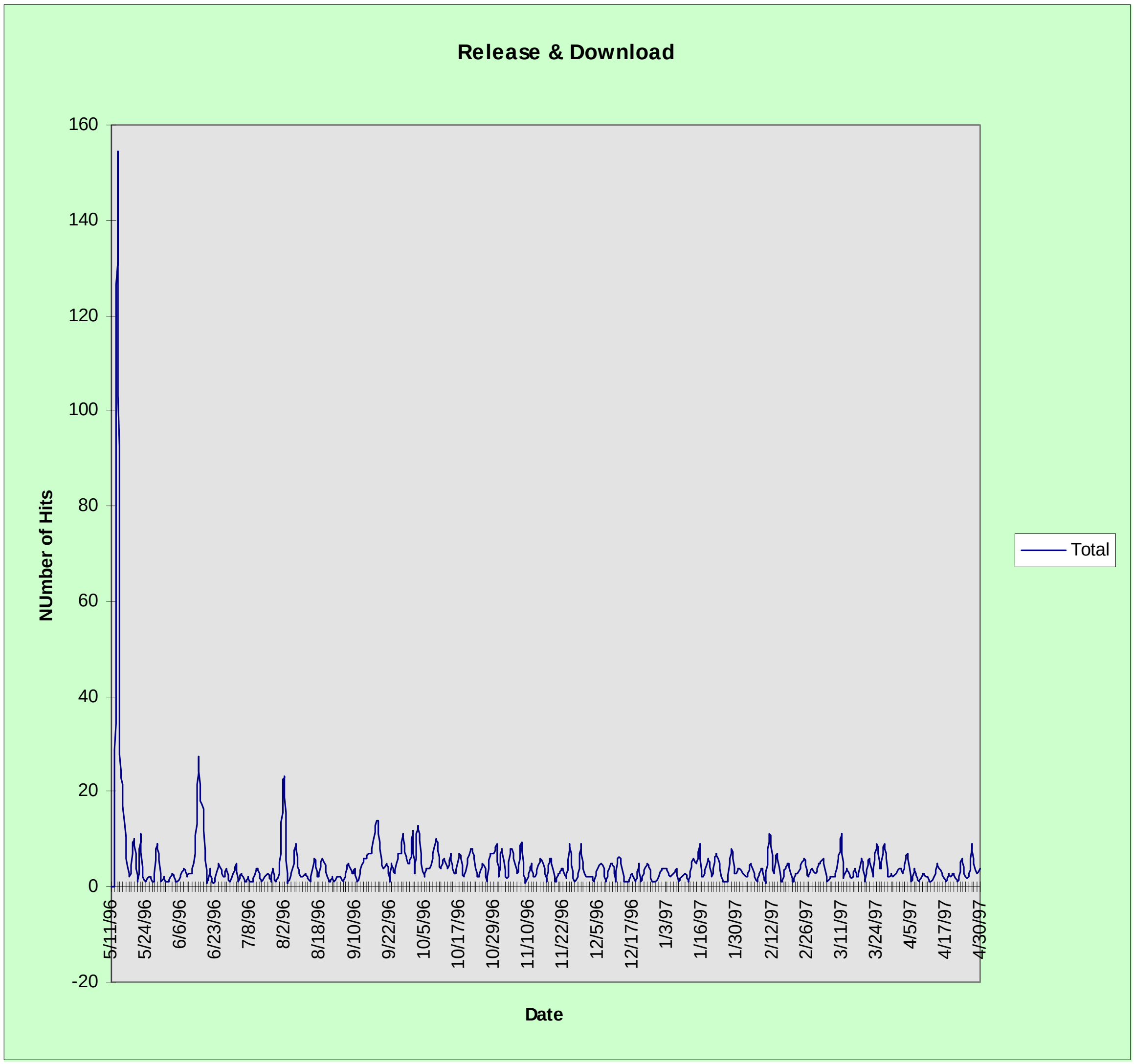


**Figure 41 CHRONO Diffusion Pattern**

We can see the first announcement on May 13, 1997 induced the highest interest and the second announcement around August 1, 1997 had a lesser impact. This pattern matches the classical *inhibitation effect* in biological stimulus-response model of neurons and animal behaviours. That is, if we apply two stimuli to a neural channel, the first (novel) stimulus has an inhibitory effect on the response of the second (follow-up) stimulus (Brown & Rothery, 1993).

As expected, the diffusion pattern after the third Yahoo announcement of September 9, 1996 is that of stochastic fluctuation because the nature of this particular dissemination channel. The only unexpected data point was that around June 15, 1996 with the second highest hit rate. As it turned out, there was a secondary dissemination effect during that period: someone associated with a SUN OS user group re-posted the first announcement to the user group listserver. In retrospect, I received several email inquiries from four users on that list about CHRONO. By examining the access_log and referer_log, we can see that many SUN sites associated with that list had set up CHRONO mechanisms around that time.

There are other statistics gathered from the log files which track CHRONO reproduction and migration patterns, for example:

- There were 7188 instances of people loading images or CGI scripts from external pages
- 1501 hits to the site came from links on external pages
- 16 came from files on the user's local computer (bookmarks)
- 0 references came from USENET news articles
- 515 came from popular web search engines (Yahoo, Lycos, Alta Vista, etc.)

| **Search Engine and Web Indexer** | **Totals** |
|---|---|
| Yahoo | 515 |
| Lycos | 0 |
| WebCrawler | 0 |
| Infoseek | 0 |
| Alta Vista | 0 |

**Table 19 Search Engines and Indexes**

Table 19 depicts incidences of memetic transmission in terms of the 5 popular web search engines and indexes.

```
/CHRONO/gif/chrono1.gif total: 85 Sites

 3133 http://www.cpcug.org/new/whatsnew.html
 2188 http://cpcug.org/new/whatsnew.html
  210 http://millkern.com/whatsnew.html
  110 http://www.cqi.com/~pmurphey/recent.html
  107 http://infobot.eng.sun.com:80/WhatsNew.html
   47 file:/export/home/sunsite/chrono/htdocs_chrono.html
   41 http://mafalda/~servinfo/SunSite/Chrono/chrono.html
   39 http://www.cpcug.org/
   34 http://sunsite.wits.ac.za/htdocs_chrono.html
   34 http://cpcug.org/
   27 http://www.eia.doe.gov/emeu/chrono/chrono.html
   22 http://www.eia.doe.gov/emeu/chrono/test-chrono.html
   19 http://cpcug.org/user/mohnkern/whatsnew.html
   17 http://www.cpcug.org/user/mohnkern/whatsnew.html
   13 http://sunsite.wits.ac.za/sunsite_index.html
   12 http://mcm-home/WhatsNew.html
   10 http://haclb01/all_changes.html
                                   …
```

**Table 20 `tw1` and `tw2` CHRONO Sites (with Accesses ≥ 10)**

```
/CHRONO/gif/chrono2.gif total: 52 Sites

  190 http://www.mohnkern.com/beloit/new.html
  139 http://petrified.cic.net/MsqlCGI/time-index.html
   95 http://www.multimedia.ncsu.edu/wnew/
   91 http://sunsite.univalle.edu.co/Chrono/Chrono.html
   63 http://www.bjp.org/whats_new.html
   50 http://www.mohnkern.com/new.html
   49 http://bjp.org/whats_new.html
   45 http://sunsite.univalle.edu.co/Chrono/chrono.html
   32 http://multimedia.csc.ncsu.edu/wnew/
   18 http://petrified.cic.net/~altitude/tmp/chrono.out2.html
   15 http://www.surfnet.nl/surfnet/Admin/stats/chrono-new.html
   13 http://www.wideopen.igc.org/amnesty/chronoaiusa.html
                                   …
```

**Table 21 `3yh` CHRONO Sites (with Accesses ≥ 10)**

Calculations based on log-files from May 13, 1996 till April 30, 1997, show there were a total of 137 CHRONO pages on the web. Table 20 shows a partial list of websites to where CHRONO mechanisms migrated. There were 85 sites with CHRONO pages of `tw1` and `tw2` lineages (i.e., www-talk). Table 21 shows there were 52 CHRONO pages of `3yh` lineage (i.e., Yahoo). Both tables show only sites that had access rates ≥ 10.

We can obtain an estimation based on statistics of the actual reproduction and of memetic transmission rates. For example, the Yahoo channel had about 10% *reproduction success rate* (52/515 = 10.09%). This reproduction success rate has been reasonably stable. For example on July 20, 1997, I compiled some follow-up measurements (e.g., numbers of sites with `3yh` lineages were 60; with `tw1` & `tw2` were 100; and Yahoo references were 660). The rate of reproduction successes through Yahoo was 60/660 = 9.09%.

### *9.4 Meta Monitor*

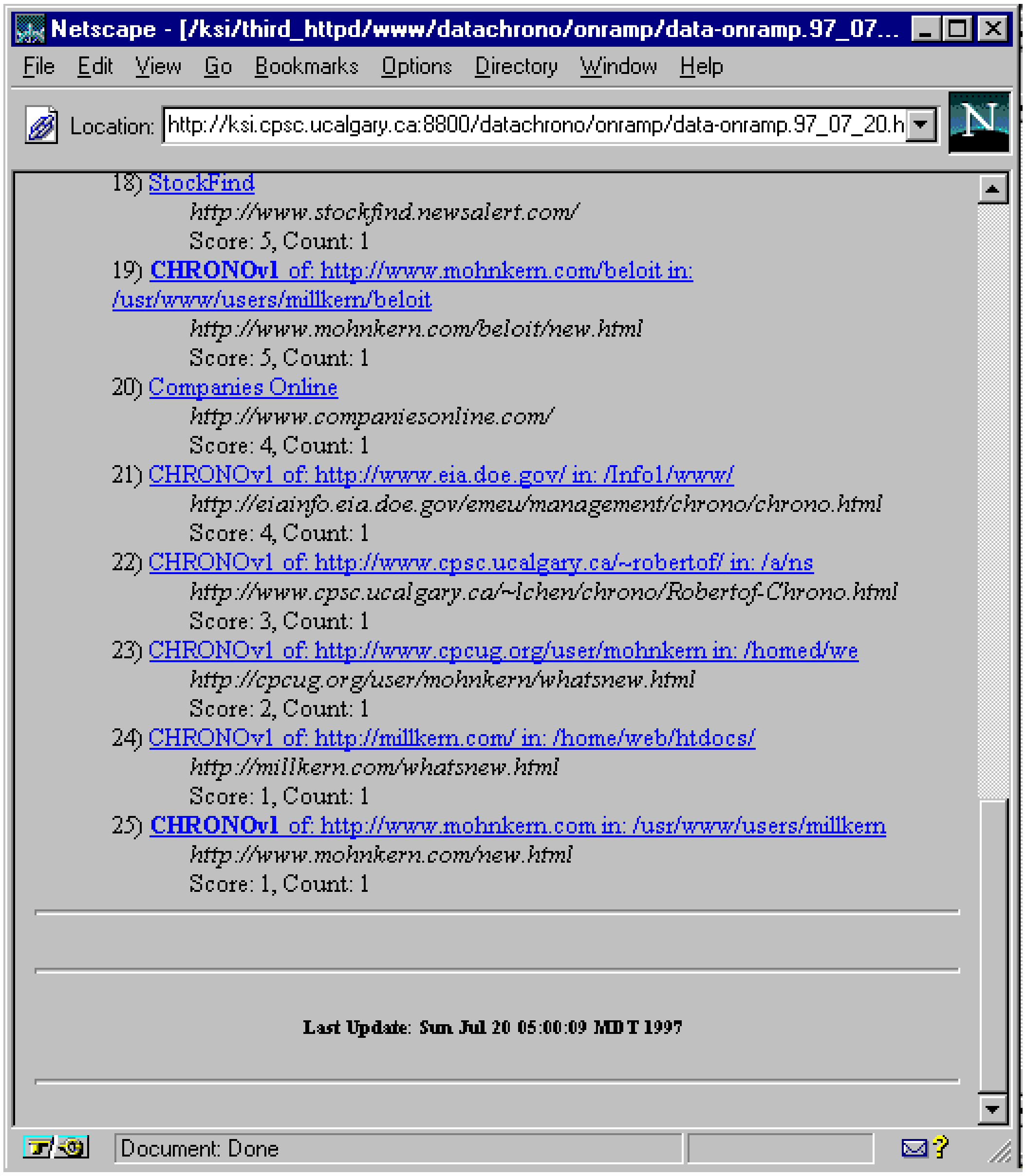


**Figure 42 Meta Monitor for the Onramp Search Engine**

Another effective channel for memetic dissemination is the search engine. Search engines like Alta Vista, WebCrawler, Lycos and Onramp are some of the more popular *resource* *awareness* mechanisms on the web. It would be useful to have a meta-level monitor that periodically sees how well a search engine is doing in picking up sites and disseminate their memes. We can use these search engines to determine how many CHRONO enabled websites are actually reachable through these search engines' robots.

The information gathered by a *meta monitor* can be pegged against the statistics from the referer_log and access_log. From those logs we can analyze CHRONO usage, find out where each CHRONO sites are, and how many sites there are. The meta monitor provides us a way to investigate the effectiveness of a search engine in web indexing. Because, from our experiment many copies of CHRONO have already been distributed through out the web, and because we know exactly how many copies are out there, we will have a good population set for testing a search engine's ability to find them.

Originally, I implemented a Perl-script meta monitor which gathers and archives daily reports of Alta Vista's[10] search results for CHRONO. It was activated on July 27, 1996. For a long period, it could only find 6 CHRONO sites. Once every few months, the meta monitor would break down due to the periodic changes of Alta Vista's output format. By November 21, 1997, Alta Vista had changed its output format once again; and it had only found 8 CHRONO sites (out of 98 sites). Its future prospect looked dismal (but it was still better than Lycos[11] which picked up only 3 to 6 sites around the same period), so I changed the monitoring focus to Onramp[12] (itself a meta-search engine that gathers results from other search engines).

I wrote a second meta monitor[13] for the Onramp meta-search engine. Since November 24, 1996, the meta monitor for Onramp has been able to discover around 12 to 35 CHRONO sites; The *success rate* of Onramp fluctuates because it depends on other search engines

---

[10] Alta Vista search engine, via http://altavista.digital.com/

[11] Lycos search engine, via http://lycos.com/

[12] Onramp search engine, via http://search.onramp.net/

[13] Onramp meta monitor archive, in http://ksi.cpsc.ucalgary.ca:8800/datachrono/onramp/

for gathering results. For example, on July 20, 1997 (Figure 42), it found 25 CHRONO sites (out of 160 sites = 100+60 lineage sites); and just a day before (July 19), it found 35.

The meta monitor technique, in conjunction with baseline statistics gathered from homing beacon and genealogy tracer techniques, provides researchers with a systematic methodology for measuring the effectiveness of dissemination channels. The underlying concept is similar to that of *sample surveying* technique (McClave & Dietrich, 1991), but instead of using sample size to estimate the underlying population, we have in this case, the underlying population with which to begin. By knowing the underlying population of a particular data set, we can evaluate the actual performance of dissemination channels.

### *9.5 Summary*

The third imperative of living systems is *propagation of the system through reproduction and/or dissemination*. Reproduction is the ultimate mechanism for genetic survival. For memes, dissemination serves a similar function. The more widespread an idea is, the more likely it will survive. This chapter presents techniques and methodologies for tracking vehicle/memetic reproduction, migration and dissemination. The focal level of investigation here is with the Internet community at large. CHRONO offers an ideal case for conducting an observational experiment for tracking software diffusion processes on the net. It is used to illustrate the *homing beacon*, *genealogy tracer*, and *meta-monitor* techniques and associated methodologies for tracking the diffusion process on the net.

# CHAPTER 10

# Conclusion

This chapter concludes the dissertation by summarizing and evaluating key ideas developed in the cyberorganism framework. The evaluation addresses the objectives described in Chapter 1 and demonstrates how these objectives are met by the research work. In addition to the evaluation, this chapter describes certain avenues of future research based on the current work.

## *10.1 Research Aim and Objectives*

The overall aim of the research reported in this dissertation is to develop a model of the socio-technical dynamics of the Internet in supporting cooperative interactions and to formulate methodologies and techniques for investigating the model. The focus here is to gain a better comprehension of the underlying dynamics involved when people collaborate through networked computing environments.

As stated in Chapter 1, The overall aim is associated with the following objectives:

1. To determine the appropriate form of model for the Internet in supporting virtual cooperative interaction.
2. To characterize the processes of cooperative interactions that have evolved on the Internet.
3. To analyze Miller's living systems theory for its application to the socio-technical cooperation through the Internet.
4. To define the socio-technical processes underlying virtual cooperative interaction.

5. To analyze the roles of awareness and its technical support in virtual cooperative interaction.
6. To develop methodologies and analysis software for modeling discourse patterns and social structure in virtual cooperative interaction.
7. To evaluate the methodologies and apply them to a sample virtual community.
8. To develop techniques and methodologies for analyzing diffusion processes in the Internet community at large.

## *10.2 Addressing the Objectives*

The research attempts to investigate the Internet from a perspective which characterizes it as an *open, dynamic system of inter-connected machines, users, and resources*. Therefore, various key ideas or *constructs* involved in networked computing environments, individual cognitive processes, and knowledge creation/dissemination processes and social systems need to be identified and explored for their inter-relationships.

For instance, the construct—*system levels*—in the *living systems theory* (Miller, 1978) provides an interesting perspective for conceptualizing the Internet. It highlights the dynamic interactions among individuals, groups, organizations and communities. From this living systems theory perspective, individuals constitute the fundamental level of analysis in the overall system. To continue with this line of reasoning: self awareness provides a sense of identity, purpose, and consciousness to each individual. When perceiving all levels collectively, the notion of self awareness can be extended to different hierarchical partitions in the system for the purpose of analyzing the inter-relationship between the overall system and its parts in the *collective stance* (Gaines, 1994b). Hence, *collective awareness* occurs at every level above the individual level in the system hierarchy (the team, special interest community and the net community at large levels). In order to function as a coherent whole, mechanisms for maintaining collective awareness in the group, organization and community levels will need to exist first. Gradually, there are some tools and services on the Internet that have emerged to fill the roles of providing awareness maintenance for individuals and groups.

The dissertation structure attempts to address the objectives in a linear fashion. Many constructs like *virtual cooperative interaction*, *memetic evolution*, and *cyberorganism* are gradually introduced through out the chapters.

### 10.2.1 Objective 1

> *To determine the appropriate form of model for the Internet in supporting virtual cooperative interaction.*

Chapter 1 introduces the *living systems theory* (Miller, 1978) as the theoretical foundation. The theory was built upon a search for the common properties of all living systems. Living systems are concrete, open systems possessing the characteristics of life. Conceptually, virtual cooperative interaction is involved with the exchange of resources between cybernetic living systems as depicted in the *systems dyad model*. Living systems, as open systems, exhibit interaction patterns with their surrounding environment and other living systems.

Chapter 2 presents three broad purposes or imperatives of life exhibited by *genes*, *memes* and their *surviving vehicles*—i.e., the host living systems: The first is immediate survival of the system through *maintenance of steady states*; the second imperative is *actualization of the system's potential* which requires both growth (i.e., incorporation of additional elements into system) and the development (i.e. elaboration of the system to cope with greater complexity in the environment); and the third imperative is *propagation of the system through reproduction and/or dissemination*. The chapter also introduces a discussion of general characteristics of all living systems.

Memes are self-replicating ideas. They share many similar characteristics with genes. The essence of memes, as with genes, is information (i.e., pattern or structure) that is capable of replicating itself. Memes are potentially immortal, although subject to mutation. As genes generate various patterns of life at the cellular, organic, and organismic levels;

memes generate a variety of cultural patterns among individuals, groups, organizations, communities and societies. Memes require survival machines or *vehicles* (Dawkins, 1989b; Hull, 1988) for their maintenance, actualization and propagation. The Internet as the global cyberorganism is their essential survival vehicle in cyberspace. Virtual cooperative interactions are the processes by which they to achieve actualization and propagate.

Chapter 2 also states the thesis that the Internet is an emergent global cyberorganism living in *cyberspace* delineated by Popper's (1972) *World 3*. The cybernetic, knowledge-creative, super-organism, called *cyberorganism*, is an aggregated living system based on *human-computer symbiosis* (Licklider, 1960). The primary purpose of the cyberorganism is to ensure the survival and propagation of memes in the cyberspace.

The cyberorganism framework attempts to model the relationship between the interaction of the individuals and their networked computing systems (i.e., human-computer interaction), and the interaction of the individuals and their virtual community (i.e., person-community interaction). This relationship creates the socio-technical dynamics underlying virtual cooperative interaction.

### 10.2.2 Objective 2

> *To characterize the processes of cooperative interactions that have evolved on the Internet.*

Chapter 2 describes the socio-technical origins and evolution of the Internet by exploring its technical origins, its actualization through development and growth, and its transformation into a global phenomenon. *Memetic mixing* and *transmission* via vehicles like *gametes* and *zygotes* are essential in its rapid development and growth. The chapter then examines the origins of *virtual cooperative interaction*. Scientific ethos, value systems, social norms are the deep roots for its motivation and reinforcement processes. The chapter concludes with suggestions for empirical research on net-based cooperative interaction.

### 10.2.3 Objective 3

*To analyze Miller's living systems theory for its application to the socio-technical cooperation through the Internet.*

Chapter 2 applies the living systems theory to classify the Internet services with a section on the physiology of the cyberorganism. In examining the functionality of the living system, it is useful to classify the major components of its subsystems in terms of the significant distinctions that determine their relative utilities. A general taxonomy of Internet services thus characterizes the major net services in terms of their utility for computer-mediated communication, access to services or search.

Chapter 3 defines 20 critical subsystems, their processes and structures in Miller's theory as they manifest themselves in human individuals and the cyberorganism. The primary structures of living systems are associated with the critical subsystems, each of which performs a function or set of vital processes for them. The cyberorganism framework identifies 20 such critical subsystems, governed by the decider subsystem. Although functionally distinct, these subsystems share components and interact closely. The cyberorganism framework focuses on key critical subsystems processing information involved in maintenance, control and coordination within the cyberorganism.

### 10.2.4 Objective 4

*To define the socio-technical processes underlying virtual cooperative interaction*.

Chapter 4 investigates the decider, timer, net/channel, associator, and memory subsystems in respect to coordination processes in the cyberorganism.

The decider subsystem is the executive centre that receives data from various sources throughout the channel and the net, then sends control information to all parts of its system. Feedback processes within a living system are needed to *maintain steady states of performance* (which is the primary imperative of life). *Awareness mechanisms* in a supra-living system transmit coordination signals among its subsystems or components. They constitute the feedback channels and regulate the adjustment processes within the supra-system.

Awareness can be viewed to be coordination signals sent among subsystems. The chapter surveys concepts and models in the CSCW research with respect to coordination and awareness processes. Collective intelligence and collective awareness can be considered cognitive models for the coordination purposes of the decider, timer, associator, and memory subsystems. Finally, this chapter examines an integrative architecture for the net/channel, associator, and memory systems in the cyberorganism.

Chapter 5 presents the five major elements for virtual cooperative interaction in the cyberorganism: (i) discourse patterns; (ii) time-dimensions; (iii) awareness hierarchy; (iv) motivations for cooperative behaviors; and (v) emergence and maintenance of virtual cooperative interaction.

The five elements model describes socio-technical processes in virtual cooperative interaction. It encompasses the communication processes and collaborative knowledge acquisition activities from closely-coupled teams to those of the very diffuse Internet community at large. It analyzes these activities in terms of the punctuated discourse processes, breaking down the cycles of action and response involved in a continuous temporal dimension. It analyzes them also in terms of awareness by originators of recipients and vice versa. The temporal dimension and awareness hierarchy enable the existing taxonomies and models of CSCW to be extended to encompass a very wide range of systems operating in both the short- and long-term and ranging from small teams to large communities. The model analyzes motivational aspects of virtual cooperative interactions. It gives rise to natural structural analysis of the activities which allows the types of communities involved to be identified from their observed activities.

### 10.2.5 Objective 5

*To analyze the roles of awareness and its technical support in virtual cooperative interaction.*

Chapter 5 develops methodological dimensions for studying and supporting awareness on the web. It describes CHRONO, an awareness maintenance mechanism for providing a feedback channel for changes at subsystems or components. The last section uses two key dimensions of the methodological framework to classify CHRONO and related mechanisms, and to clarify the human factors design issues involved.

There are four main dimensions of design considerations for awareness maintenance artifacts for web users:

1. Locus of Responsibility: *Server-Side*, *Client-Side*, or *Centralized Dispatcher*
2. Level of Awareness Hierarchy: *Team*, *Special Interest Community*, or *Internet Community at Large*.
3. Method of Locating Changes: *Browsing* vs. *Targeting*
4. Complexity of User Interaction: *Simplicity* vs. *Customization*

The first dimension, the *locus of responsibility*, differentiates who is responsible for maintaining the record-keeping mechanisms for supporting awareness maintenance. The second dimension, the *level of awareness hierarchy*, reflects the need for maintaining mutual awareness among members existing in various collaborative arrangements, and there are three main levels of awareness arrangements which constitute the awareness maintenance hierarchy. The third dimension, the *method of locating changes*, involves two different ways of locating documents that have been changed: browsing and targeting. Finally, the fourth dimension, the *complexity of user interaction*, denotes system usability in terms of simplicity vs. customization.

The chapter concludes by presenting a classification scheme using the first two key dimensions of the methodological framework to classify CHRONO and related mechanisms, and to clarify the human factors design issues involved.

### 10.2.6 Objective 6

*To develop methodologies and analysis software for modeling discourse patterns and social structure in virtual cooperative interaction.*

Chapter 7 proposes systematic methodologies for analyzing listservers on the net. One of the key objectives is to gather various types of useful statistics for gauging the life cycle and vitality of virtual communities on the net through observable interactions at the special interest community level in the cyberorganism.

A better understanding of discourse processes in special interest communities can be obtained by conducting social-network analyses (Wellman & Berkowitz, 1988) and by using the SYMLOG observational approach (Bales & Cohen, 1979; Losada, Sanchez, & Noble, 1990) adjunct with time-series analysis (Suen & Ary, 1989) on communication patterns on listservers.

To analyze the social network of a special interest community, three types of measurements are useful in summarizing list members' interactions: time-series plot, sociogram, and SYMLOG field diagram. The chapter introduces an overall coding scheme with key fields, such as: who posts to the list, who replies to whom, or which topic thread is growing. The coding scheme is used to produce various coding tables. Many useful descriptive statistics then can be determined from the resulting coding tables.

Measurements like indegree, outdegree, degree centrality, and degree prestige can be calculated for identifying important or influential members in the social network of a list. For content analysis, SYMLOG field diagrams can be plotted after sequential event-coding of posted messages. In conjunction, sociograms and SYMLOG field diagrams allow us to visualize the group-dynamics of special interest communities.

Chapter 8 introduces ListA: a Listserver Analysis program and a sample case special interest community—the Global Brain Community—for demonstrations of ListA and the systematic methodologies described in Chapter 7.

ListA consists of 5 analysis tools: (1) PREP: preparation of data files; (2) TIME SERIES: time-series plot; (3) SOCIOGRAM: generation of dynamic interactive sociogram; (4) ENCODE SYMLOG: for rating SYMLOG interaction scores; (5) FIELD DIAGRAM: group average field diagrams (original & adjusted).

In terms of content-free analysis, ListAnalyzer can automatically produce data files with sorted key fields, a time-series plot and associated descriptive statistics, a sorted sociomatrix, and an interactive sociogram. Together, they can give fast feedback for researchers or/and list members about a special interest community's vitality, social structure, and life cycle. From the resulting feedback, researchers can determine whether or not the further content analysis of the community is worthwhile. In order to analyze the socio-psychological dynamics of the community, social scientists can use content analysis using the SYMLOG methodology. ListA provides efficient coding of observed interaction following the SYMLOG formalism and generates Group Average Field Diagrams (original and adjusted) once the coding phase has been completed.

### 10.2.7 Objective 7

*To evaluate the methodologies and apply them to a sample virtual community.*

Chapter 8 chooses the Global Brain special interest community as a test case for analyses of its life cycle, evolution, social structure, members' contribution, centrality, influence, leadership and group dynamics.

The ListA software is used to prepare data files for analyzing the Global Brain list. It generates a time-series plot for analyzing the list's life cycle, and a dynamic interactive sociogram for investigation of its social networks. In addition, ListA offers input-facilitation of SYMLOG coding schemes in respect to individual postings. After the coding phase has been completed, ListA allows researchers interactively to view SYMLOG field diagrams of Global Brain list for further analysis.

In essence, the chapter uses the Global Brain list to evaluate the effectiveness of the systematic methodologies presented earlier in Chapter 7 and the associated analysis software in Chapter 8.

### 10.2.8 Objective 8

> *To develop techniques and methodologies for analyzing diffusion processes in the Internet community at large.*

Chapter 9 introduces techniques and methodologies for tracking diffusion processes on the net. Reproduction is the ultimate mechanism for genetic survival. For memes, dissemination serves a similar function. The more widespread an idea is, the more likely it will survive.

This chapter presents techniques and methodologies for tracking vehicle/memetic reproduction, migration and dissemination. The focal level of investigation here is at the Internet community at large. CHRONO offers an ideal case for conducting an observational experiment for tracking software diffusion processes on the net. Homing beacon, genealogy tracer, and meta-monitor are three techniques for three techniques introduced here:

- The *homing beacon* technique of embedding a tracking GIF icon in software is used for software usage tracking methodology. The GIF should be strategically placed (e.g., at the bottom of the page), so that users do not need to wait for its fetching process and will be able to read the page without inconvenience.
- The *genealogy tracer* technique of using a CGI Release Page for tagging, together with the systematic strategy of selective announcement, provides us a methodology to investigate the effectiveness of different diffusion channels and to analyze the resulting pattern of memetic awareness in the cyberorganism.
- The *meta monitor* technique for monitoring dissemination channels, in conjunction with baseline statistics gathered from homing beacon and genealogy tracer techniques,

provides researchers a systematic methodology for measuring the effectiveness of dissemination channels.

These three techniques and their associated methodologies offer researchers useful tools for analyzing reproduction, migration and dissemination processes in the global cyberorganism

## *10.3 Future Work*

This section describes potential future research directions based upon the cyberorganism framework beyond its present scope.

### 10.3.1 Analysis and Development of New Net Services

The purpose of the research reported in this dissertation has been to develop a finer-grained conceptual model for communication, knowledge and social processes that occur in the global cyberorganism. In order to support and improve those processes through new and better services, the cyberorganism framework suggests three levels of analysis of services:

- *Message quality*—the improvement of the multimedia capabilities of the basic message channel—there has been continuous improvement from simple text to typography, images, movies, sounds, animations, simulations, and so on.
- *Relationship modeling*—the incorporation of linkage information preserving discourse relationships—the hypertext links of the original web technology introduced this capability and clickable maps extended it—there is scope for further extension based on greater understanding of the roles that the links play in enabling people to grasp the argument forms of information on the web.
- *Awareness support*—the systematic reduction of the time (t2 and t4 in Figure 12) for a potential recipient to become aware of relevant information—manual and automatic indexing and various forms of search engines have made massive advances in coping with the information overload resulting from the growth of the web—however, there

is scope for many different tools supporting the various ways in which people manage their awareness.

The key question to ask in developing new awareness support mechanisms is "what is the starting point for the person seeking information, the existing information that is the basis for the search." A support mechanism is then one that takes that existing information and uses it to present further information that is likely to be relevant. Such information may include relevant concepts, text, existing documents, people, sites, listservers, newsgroups, and so on. The support mechanism may provide links to further examples of all of these based on content, categorization or linguistic or logical inference. The outcome of the search may be access to a document, but it may also be email to a person, a list or a newsgroup (Chen & Gaines, 1997c).

By considering the net as a whole in the cyberorganism framework, in the future I will attempt to model and track the evolution and ecology of net services as they emerge to fill functional niches, very much like chronological awareness mechanisms have done so.

### 10.3.2 Auto Referrer Mechanism

As noted earlier in the dissertation, one aspect of the web is the asymmetric awareness between information resource providers and information resource users. For example, a person may maintain a website devoted to literature surveys in population biology. Such a specialty website tends to attract other people who are interested learning about population biology. It would be mutually beneficial if they could exchange ideas about their shared interests. A site visitor has the option to initiate the contact through email service.  However, the social exchange theory dictates there is a social cost involved in establishing explicit contact and exchange. In order to be able to exchange ideas freely, there has to be an establishment of a 'relationship' between two parties.

Some websites have a *guest book* and ask visitors to sign it and write down comments or/and a hyperlink to their websites. The motivation behind this is that the website maintainer may like to know about people who have similar interests. However, this strategy also requires visitors to take the time and make the effort to sign the guest book. Some people will not bother to do so.

Nevertheless, if a visitor thinks that the information resource is worth having his own page linked to it, this certainly is highly indicative that he and the information resource provider already share similar interests. It would be advantageous for the resource provider to be able to have an *automatic referrer mechanism* (REFERRER) to determine which pages have actually *linked* to her site. She then follows *reverse hyperlinks* to those web pages.

The basic infrastructure for such a resource awareness mechanism has already been discussed in Chapter 9 (tracking memetic diffusion) and Chapter 6 (CHRONO mechanism.)

Since a referer_log already records the page that refers to any page on a particular website, we can set up a *cron process* very much like that of CHRONO. It periodically goes through the referer_log and extracts the http links to particular web pages that a site maintainer wants to monitor.

A HTML tag like:

```
<meta auto-referrer>
```

can be placed in the desired pages which flags REFERRER to compile hyperlink indexes for them. REFERRER then automatically substitutes a specific hyperlink icon to an associated index page per desired page:

```
<meta auto-referrer>
<A HREF=" http://ksi.cpsc.ucalgary.ca:8800/referrer/78909323.html">
   <IMG SRC="http://ksi.cpsc.ucalgary.ca:8800/referrer/referrer.gif">
</A>
```

Afterwards, people can click on the REFERRER Icon on the tracked pages to jump to other websites with similar interests.

The auto referrer mechanism described above is currently under investigation for future development. It shares the same spirit as *collaborative social filtering* (Maltz & Ehrlich, 1995; Shardanand & Maes, 1995) and *recommender* (Schwartz & Wood, 1993; Resnick & Varian, 1997; Resnick, 1997) systems which recommend useful information (e.g., video or music picks) through association mechanisms. They rely on the principle that people who share similar interests may cooperate virtually, without direct extensional awareness of one another.

### *10.4 Summary and Conclusion*

The 1990s have seen the emergence of large scale cooperative activities on the Internet using email, listservers, newsgroups and the World Wide Web. There have also been developments of systems using some of these technologies to support smaller closely-coupled teams. In terms of the standard time/space taxonomy for CSCW, these uses of the Internet are generally virtual in space and range from highly synchronous to highly asynchronous interactions. However, many of the major applications of the Internet raise new issues that are not adequately addressed by existing models and taxonomies of CSCW.

Small groups of individuals working together generally have well-defined roles and mutual awareness of roles, tasks and activities. However, on a listserver, a discussion may be initiated with only a vague concept of other potential participants but with strong expectations that a collaborative activity will result. On the web, material may be published with only a vague conception of potential users, yet that material may play an essential role in a collaborative active in some community, possibility not involving the originator, and perhaps a community of which the originator is not part. These phenomena are common in various collaborative scientific communities conducting interdisciplinary research. Those loosely collaborative virtual communities are moving their knowledge acquisition processes to the Internet and the web. The net as a global cyberorganism can be considered as a large-scale groupware for supporting special interest communities (e.g., high-energy physics research community). Large scale groupware for virtual communities differs not only in the quantity, but also in the quality of cooperative interaction (Dennis, Valacich & Nunamaker, 1990). It would be interesting to know whether they can be modeled and supported using some extended CSCW frameworks.

The net is a vehicle for discourse in which the goals of individual agents are supported through social knowledge processes, and support tool design needs to be based on increasingly refined models of those processes. Much of the current research is concerned with the empirical studies of discourse processes on the net through analysis of information diffusion, listserver archives, and so on. Tools which develop models of such processes are useful for social scientists and CSCW researchers. Making them available to the participants, may result in improved usage of net resources.

A major contribution of current research is the development of a new conceptual framework for modeling socio-technical dynamics on the Internet. This dissertation contributes to CSCW research by drawing attention to the significance of virtual cooperative interaction in computing networks such as the web where social and organizational structures are fluid and less defined.

The cyberorganism framework for virtual cooperative interaction expands the scope of groupware research. It provides a conceptual framework encompassing all forms of distributed knowledge creation and dissemination processes from teams through special interest community to diffused, evolving global cyberorganism. Modeling and supporting virtual cooperative interaction on the Internet are important new challenges for research in computer-mediated communication, human-computer interaction and computer-supported cooperative work.

Finally, the dissertation offers investigative techniques and research methodologies for the studying of social-technical processes. They are valuable in assisting CSCW researchers and social scientists for further empirical investigations of the socio-technical processes on the Internet.

# References


Andreessen, M. (1993). *NCSA Mosaic Technical Summary*. NCSA, University of Illinois. ftp://ncsa.uiuc.edu/Web/Mosaic/Papers/mosaic.ps.Z

Baecker, R. M. [ed.] (1993). *Readings in Groupware and Computer-Supported Cooperative Work: assisting human-human collaboration*. San Mateo, CA: Morgan Kaufmann.

Baecker, R. M., Nastos, D., Posner, I. R. and Mawby, K. L. (1993). The User-Centered Iterative Design of Collaborative Writing Software. In *Proceedings of ACM INTERCHI '93*, pp. 399-305.

Bailey, K. D. (1994). *Sociology and the New Systems Theory: toward a theoretical synthesis*. Albany, NY: SUNY Press.

Bakeman, R. and Gottman, J. M. (1986). *Observing Interaction: an introduction to sequential analysis*. UK: Cambridge University Press.

Bakeman, R. and Quera, V. (1995). *Analyzing Interaction: sequential analysis with SDIS and GSEQ*. Cambridge, UK: Cambridge University Press.

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

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

Bandura, A. and Walters, R. (1963). *Social Learning and Personality Development*. New York, NY: Holt, Rinehart & Winston.

Baran, P. et al. (1964). On Distributed Communications, Vols. I-XI. RAND Corporation Memos, August, 1964.

Bayers, C. (1996). The Great Web Wipeout. *Wired*, Vol. 4, No. 4, pp. 126-128.

Benedikt, M. [ed.] (1991). CyberSpace: First Steps. Cambridge, MA: MIT Press.

Berners-Lee, T. (1993). World-Wide Web Talk at Online Publishing 1993. CERN, Geneva. http://info.cern.ch/hypertext/WWW/Talks/OnlinePublishing93/Overview.html.

Berners-Lee, T. (1997). World-Wide Computer. *Communications of the ACM*, February, Vol. 40, No. 2, pp. 57-58.

Berners-Lee, T. and Cailliau, R. (1990). *World Wide Web: Proposal for a Hypertext Project*. CERN, Geneva. http://info.cern.ch/hypertext/WWW/Proposal.html

Berners-Lee, T., Cailliau, R., Luotonen, A., Nielsen, H. F. and Secret, A. (1994). The World-Wide Web. *Communications of the ACM*, August, Vol. 37, No. 8, pp. 76-83.

Blume, S. S. (1974). *Toward a Political Sociology of Science*. New York, NY: The Free Press.

Bowman, C. M., Danzing, P. B., Manber, U. and Schwartz, M. F. (1994). Scalable Internet Resource Discovery: research problems and approaches. *Communications of the ACM*, August, Vol. 37, No. 8, pp. 98-107.

Boyd, R. and Richerson, P. J. (1985). *Culture and the Evolutionary Process.* Chicago, IL: University of Chicago Press.

Brothers, L., Hollan, J., Nielsen, J., Stornetta, S., Abney, S., Furnas, G., and Littman, M. (1992). Supporting Informational Communication via Ephemeral Interest Groups. *Proceedings of CSCW 92*, ACM Press, pp. 84-90.

Brown, D. and Rothery. P. (1993). *Models in Biology: mathematics, statistics and computing*. New York, NY: John Wiley & Sons.

Bruckman, A. (1994). Programming for Fun: MUDs as a context for collaborative learning. Submitted to *NECC '94*.

Bruckman, A. and Resnick, M. (1993). Virtual Professional Community: results from the MediaMOO project. *The Third International Conference on Cyberspace*, Austin, TX, May 1. ftp://media.mit.edu/pub/MediaMOO/papers/

Bush, V. (1945). As We May Think. From *The Atlantic Monthly*, July 1945, pp. 101-108. Reprint in: Goldberg, A. [ed.] (1988), pp. 237-247.

Capra, F. (1996). *The Web of Life: a new scientific understanding of living systems*. New York, NY: Anchor Books.

CERN (1994). *History to Date*. CERN, Geneva.
http://info.cern.ch/hypertext/WWW/History.html

Chen, L. L.-J. (1996). Chronological Awareness Tools: CHRONO and Meta-CHRONO. *Proceedings of KAW96: The 10$^{th}$ Knowledge Acquisition Workshop*, Banff, Alberta, November 9-14, 1996,
http://ksi.cpsc.ucalgary.ca/KAW/KAW96/chen/kawchrono.html

Chen, L. L.-J. and Gaines, B. R. (1996a). Knowledge Acquisition Processes in Internet Communities. *Proceedings of KAW96: The 10$^{th}$ Knowledge Acquisition Workshop*, Banff, Alberta, November 9-14, 1996, pp. 43-1–43-18.

Chen, L. L.-J. and Gaines, B. R. (1996b). Methodological Issues in Studying and Supporting Awareness on the World Wide Web. *Proceedings* of the *WebNet96 Conference*, San Francisco, CA., Oct. 16-19, Association of Advancement of Computing in Education (AACE), pp. 95-102.

Chen, L. L.-J. and Gaines, B. R. (1997a). A CyberOrganism Model for Awareness in Collaborative Communities on the Internet. *International Journal of Intelligent Systems (IJIS)*, Vol. 12, No. 1. pp. 31-56.

Chen, L. L.-J. and Gaines, B. R. (1997b). Awareness and Virtual Cooperative Interaction in the Learning Web. *Proceedings of EdMedia/EdTelecom '97 Conference*, Calgary, Alberta, June 14-19, 1997. Association of Advancement of Computing in Education (AACE).

Chen, L. L.-J. and Gaines, B. R. (1997c). Modeling and Supporting Virtual Cooperative Interaction Through the Web. Book chapter in: Sudweeks, F. McLaughlin M. L. and Rafaeli. S., [eds.] (1997).

CommerceNet (1995). *CommerceNet/Nielsen Internet Demographics Survey*. CommerceNet. http://www.commerce.net/information/surveys/

Cook, K. S., [ed.] (1987). *Social Exchange Theory*. Newbury Park, CA: Sage.

Crane, D. (1972). *Invisible College: diffusion of knowledge in scientific communities*. Chicago, IL: University of Chicago Press.

Crocker, S. (1969) Host Software. *RFC 001*, April 7, 1969.

Csikszentmihalyi, M. (1990). *Flow: the psychology of optimal experience*. New York, NY: Harper and Row.

Csikszentmihalyi, M. (1993). *The Evolving Self: a psychology for the third millennium*. New York, NY: Harper Collins.

Curtis, P. and Nichols, D. A. (1993). *MUDs grow up: social virtual reality in the real world*. XeroxParc Internal Report, Jan. 19, 1993.

Danet, B., Ruedenberg, L. and Rosenbaum-Tamari, Y. (1997). In: Sudweeks, F. McLaughlin M. L. and Rafaeli, S., [eds.] (1997).

Dawkins, R. (1986). *The Blind Watch Maker*, (republished in 1991). London, UK: Penguin Book.

Dawkins, R. (1989a). *The Extended Phenotype*. New York, NY: Oxford University Press.

Dawkins, R. (1989b). *The Selfish Gene, new edition* (originally published in 1976). Oxford, UK: Oxford University Press.

Dawkins, R. (1995). *River Out of Eden: a Darwinian view of life.* New York, NY: Basic Book.

Dennett, D. C. (1995). *Darwin's Dangerous Ideas: evolution and the meaning of life.* New York, NY: Touchstone Book.

Dennis, A. R., Valacich, J. S. and Nunamaker, J. F. Jr. (1990). An Experimental Investigation of the Effects of Group Size in an Electronic Meeting Environment. *IEEE Transactions on Systems, Man, and Cybernetics*, 20, pp. 1049-1059.

Dertouzos, M. L. (1997).*What Will Be: how the new world of information will change our lives*. New York: HarperCollins.

Dourish, P. and Bellotti, V. (1992). Awareness and Coordination in Shared Workspaces. *Proceedings of CSCW 92*, ACM Press, pp. 107-114.

Dunbar, R. I. M. (1988). The Evolutionary Implications of Social Behavior. In: Plotkin, H. C. [ed.] (1988), pp. 165-188.

EIT (1994). *Hypermail*. Enterprise Integration Technologies. http://www.eit.com/

Ellis, C. and Wainer, J. (1994). A Conceptual Model of Groupware. *Proceedings of CSCW 94,* ACM Press, pp. 79-88.

Engelbart, D. The Augmented Knowledge Workshop. In: Goldberg, A. [ed.] (1988), pp. 187-231.

Evard, R. (1993). Collaborative Networked Communication: MUDs as systems tools. *Proceedings of LISA*, Monterey, CA., Nov. 1-5. ftp://parcftp.xerox.com/pub/MOO/papers/Evard.ps

Flores-Mendez, R. A. (1997). Java Concept Maps for the Learning Web. *Proceedings of EdMedia/EdTelecom '97 Conference*, Calgary, Alberta, June 14-19, 1997. Association of Advancement of Computing in Education (AACE).

Gaffin A. and Heitkötter, J. (1994). *EFF's (Extended) Guide to the Internet.* http://www.eff.org/papers/eegtti/eeg_44.html

Gaines, B. R. (1971). Through a Teleprinter darkly. *Behavioural Technology*. Vol. 1, No. 2, pp. 15-16.

Gaines, B. R. (1989). Social and Cognitive Processes in Knowledge Acquisition. *Knowledge Acquisition*, Vol. 1, pp. 39-58.

Gaines, B. R. (1993a). An Agenda for Digital Journals: the socio-technological infrastructure of knowledge dissemination. *Journal of Organizational Computing*, Vol. 3, No. 2. pp. 135-193.

Gaines, B. R. (1993b). Representation, discourse, logic and truth: situating knowledge technology. Mineau, G. W., Moulin, B. and Sowa, J. F., [Ed]. Conceptual Graphs for Knowledge Representation. pp. 36-63. New York: Springer.

Gaines, B. R. (1994a). Supporting Collaboration Through Multimedia Digital Document Archives. *Proceedings of Canadian Multi-media Conference (CMMC '94*). Calgary, Alberta.

Gaines, B. R. (1994b). The Collective Stance in Modeling Expertise in Individuals and Organizations. *International Journal of Expert Systems.* Vol. 7, No. 1, pp. 21-51.

Gaines, B. R. (1996). The Emergence of Knowledge through Modeling and Management Processes in Societies of Adaptive Agents. *Proceedings of KAW96: The 10$^{th}$*

*Knowledge Acquisition Workshop*, Banff, Alberta, November 9-14, 1996, pp. 24:1-24:13.

Gaines, B. R. and Chen, L. L.-J. (1996). A Model of Knowledge Processes in Internet Communities. *Proceedings of AAAI-96 Workshop: Internet-Based Information Systems*, Portland, Oregon, August 5, 1996, American Association of Artificial Intelligence (AAAI), pp. 41-47.

Gaines, B. R. and Shaw., M. L. G. (1995). WebMap: Concept Mapping on the Web. *Proceedings of the 4th International WWW Conference*, Boston, MA.

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 of the American Society for Information Science (JASIS),* [in press].

Gaines, B. R., Shaw, M. L. G and Chen, L. L.-J. (1996). Utility, Usability and Likeability: Dimensions of the Net and Web. *Proceedings* of the *WebNet96 Conference*, San Francisco, CA., Oct. 16-19, 1996, Association of Advancement of Computing in Education (AACE), pp. 167-173.

Galegher, J. (1990). Intellectual Teamwork and Information Technology: the role of information systems in collaborative intellectual work. In: Carroll, J. S. [ed.] *Applied social psychology and organizational settings*., Hillsdale, NJ: Lawrence Erlbaum Associates.

Galegher, J. and Kraut, R. E. (1990). Computer-Mediated Communication for Intellectual Teamwork: a field experiment in group writing. In: *Proceedings of CSCW '90*, pp. 65-78.

Galegher, J., Kraut, R. E. and Egido, C., [eds.] (1990). *Intellectual Teamwork: social and technological foundations of cooperative work.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Gbrain (1996). *The Global Brain Group*. http://pespmc1.vub.ac.be/GBRAIN-L.html

Gibson, W. (1984). *Neuromancer*. New York, NY: Ace Books.

Goldberg, A. [ed.] (1998). A History of Personal Workstations. New Work, NY: ACM Press.

Goldfarb, C.F. (1990). *The SGML Handbook.* Oxford, UK: Clarendon Press.

Gotelli, N. J. (1995). *A Primer of Ecology.* Sunderland, MA: Sinauer Associates.

Gould, S. J. (1989). *Wonderful Life: the burgess shale and the nature of history*. New York, NY: W. W. Norton & Co.

Gould, S. J. (1994). The Evolution of Life on the Earth. *Scientific American,* October, Vol. 271, No. 4, pp. 84-91.

Gutwin, C., Roseman, M. and Greenberg, S. (1996). A Usability Study of Awareness Widgets in a Shared Workspace Groupware System. *Proceedings of Computer Supported Collaborative Work (CSCW 96)*, Boston, MA. Nov. 16-20. pp. 258-267.

Gutwin, C., Stark, G. and Greenberg, S. (1995). Support for Workspace Awareness in Educational Groupware. *Proceedings of the ACM Conference on Computer Supported Collaborative Learning*, Bloomington, IN., Oct. 17-20.

Hare, A. P. (1989). New Field Theory: SYMLOG research, 1960-1988. In: Lawler, E. J. and Markovsky, B. [eds.] *Advances in group processes: a research annual,* vol. 6. JAI Press, Greenwich, CN, pp. 229-257.

Harnad, S. (1994). Publicly Retrievable FTP Archives for Esoteric Science and Scholarship: a subversive proposal. *From: USENET discussions*, archived by *harnad@princeton.edu*.

Harnad, S. (1995). Implementing Peer Review on the Net: Scientific Quality Control in Scholarly Electronic Journals. In: Peek, R. and Newby, G. [eds.] *Electronic publishing confronts academia: the agenda for the year 2000*. Cambridge MA: MIT Press.

Hastings, A. (1997). *Population Biology: concepts and models*. New York, UK: Springer-Verlag.

Hearst, M. A. (1997). Interfaces for Searching the Web. *Scientific American*, March, Vol. 276 No. 3, pp. 68-73.

Hill, W. and Terveen, L. (1996). Using Frequency-of-mention in Public Conversations for Social Filtering. *Proceedings of CSCW '94*, ACM Press, pp. 106-121

Hill, W., Stead, L., Rosenstein, M. and Furnas, G. (1995). Recommending and Evaluating Choices in a Virtual Community of Use. *Proceedings of CHI '95*, ACM Press, pp. 194-201.

Hoagland, M. and Dodson, B. (1995). *The Way Life Works*. New York, NY: Times Books.

Hoffman, D. L. and Novak, T. P. (1995). *Marketing in Hypermedia Computer-Mediated Environments: conceptual foundations*. Working Paper No. 1, Research Program on Marketing in CME. Owen Graduate School, Vanderbilt University.

Hull, D. L. (1988). Interactors versus Vehicles. In: Plotkin, H. C. [ed.] (1988), pp. 19-50.

Hutchins (1990). The Technology of Team Navigation. In Galegher, J., Kraut, R. E. and Egido, C., [eds.] (1990), pp. 407-428.

Ishii, H. and Kobayashi, M. (1992) ClearBoard: a seamless medium for shared drawing and conversion with eye contact. *Proceedings, CHI '92*. ACM Press, pp. 525-532.

Ishii, H. and Miyake, N. (1991). Toward an Open Shared Workspace. *Communications ACM.* 34(12) 37-50.

Jones, E. E. and Pittman, T. S. (1982). Toward a General Theory of Strategic Self-Presentation. In Suls, J. [ed.], *Psychological Perspectives on the Self*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Jones, S. (1997). Using the News: an examination of the value and use of news sources in CMC. In: Sudweeks, F. McLaughlin M. L. and Rafaeli, S., [eds.] (1997).

Jones, S. G. (1995). Understanding Community in the information age. Jones, S. G. [ed.] *CyberSociety: Computer-Mediated Communication and Community*. pp. 10-35. Sage, Thousand Oaks, CA: Sage.

Kauffman, S. A. (1993). *The Origins of Order: self-organization and selection in evolution*. Oxford, UK: Oxford University Press.

Kauffman, S. A. (1995). *At Home in the Universe: the search for the laws of self-organization and complexity*. Oxford, UK: Oxford University Press.

Kautz, H., Selman, B. and Shah, M. (1997). Referral Web: combining social networks and collaborative filtering. *Communications of the ACM*, March, Vol. 40, No. 3, pp. 63-65.

Kelly, H. H. and Thibaut, J. W. (1978). *Interpersonal Relations: a theory of interdependence*. New York, NY: Wiley.

King, S. A. (1996). Researching Internet Communities: proposed ethical guidelines for reporting results. *The Information Society*, Vol. 13, pp. 119-127.

Kling, R. (1980). Social Analysis of Computing: theoretical perspectives in recent empirical research. *Computing Surveys*, Vol. 12, No. 1. ACM Press, pp. 61-110.

Kollock, P. (1997). Design Principles for Online Communities. In: *The Internet and Society: Harvard Conference Proceedings*. Cambridge, MA: O'Reilly & Associates.[Forthcoming]
http://www.sscnet.ucla.edu/soc/faculty/kollock/papers/design.htm

Krol, E. (1993). FYI on "What is the Internet?". *RFC 1462*, Internet.

Kuhn, T. (1962). *The Structure of Scientific Revolutions*, (republished as 2/E in 1970). Chicago, IL: University of Chicago Press.

Landow, G. P. (1990). Hypertext and Collaborative Work: the example of Intermedia. In. Galegher, J., Kraut, R. E. and Egido, C., [eds.] (1990), pp. 407-428.

Leiner, B. M., Cerf, V. G., Clark, D. D., Kahn, R. E., Kleinrock, L., Lynch, D. C., Postel, J., Roberts, L. G. and Wolff, S. (1997). The Past and Future History of the Internet. *Communications of the ACM*, February, Vol. 40, No. 2, pp. 102-108.

Lewin, K. (1951). *Field Theory in Social Science*. New York, NY: Harper.

Licklider, J. C. R. (1960). Man-Computer Symbiosis. From *IRE Transactions on Human Factors in Electronics*, March 1960, pp. 4-11. IEEE. Reprint in: Goldberg, A. [ed.] (1988), pp. 131-140.

Littlejohn, S. W. (1992). *Theories of Human Communication*, 4/E., Belmont, CA: Wadsworth.

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

Losada, M., Sanchez, P. and Noble, E. E. (1990). Collaborative Technology and Group Process Feedback: their impact on interactive sequence in meetings. *Proceedings of CSCW '90*, ACM Press, pp. 53-64.

Lottor, M. (1997). Internet Domain Survey. Network Wizards. http://www.nw.com

Kremer, R. (1997). Multi-user Interactive Concept Maps for the Learning Web. *Proceedings of EdMedia/EdTelecom '97 Conference*, Calgary, Alberta, June 14-19, 1997. Association of Advancement of Computing in Education (AACE).

Lycos (1995). Lycos. http://www.lycos.com

Lynch, C. (1997). Searching the Internet. *Scientific American*, March, Vol. 276 No. 3, pp. 52-56.

Mabry, E. A. (1997). Framing Flames: the structure of argumentative messages on the net. In: Sudweeks, F. McLaughlin M. L. and Rafaeli, S., [eds.] (1997).

Macionis, J. J. (1989). *Sociology*, 2/E. Englewood Cliffs, NJ: Prentice-Hall.

MacKinnon, R. (1997). Virtual Rape. In: Sudweeks, F. McLaughlin M. L. and Rafaeli, S., [eds.] (1997).

Maltz, D. and Ehrlich, K. (1995). Pointing the Way: Active Collaborative Filtering. *Proceedings of CHI '95*, ACM Press, pp. 202-209.

Mandviwalla, M. and Olfman, L. (1994). What Do Groups Needs? a proposed set of generic groupware requirements. *ACM Transactions on Computer-Human Interaction*, September, Vol. 1, No. 2, pp. 245-268.

Margulis, L. (1981). *Symbiosis in Cell Evolution*. San Francisco: Freeman.

Maynard Smith, J. (1993). *The Theory of Evolution*, 3/E (originally published in 1958). New York, NY: Cambridge University Press Syndicate.

McClave, J. T. and Dietrich, F. H. II (1991). *Statistics (4/E)*. New York, NY: Macmillan.

McGrath, J. E. (1990). Time Matters in Groups. Galegher, J., Kraut, R. E. and Egido, C., [eds.], pp. 23-61.

McLaughlin, M. L. (1984). Conversation: How Talk is Organized. *Sage Series in Interpersonal Communication*, Vol. 3. Beverly Hills, CA: Sage.

McLuhan, M. (1964). *Understanding Media: the extensions of man*, (republished in 1994), Cambridge, MA: MIT Press.

McLuhan, M. and McLuhan, E. (1988). *Laws of Media: the new science*. Toronto, ON: University of Toronto Press.

Merton, R. K. (1942). Science and Technology in a Democratic Order, *Journal of Legal and Political Sociology* 1, pp. 115-26. Reprint in: Merton, R. K. (1973), pp. 267-285.

Merton, R. K. (1957). Priorities in Scientific Discovery. American Sociological Review, 22. No. 6 (December 1957), pp. 635-669. Reprint in: Merton, R. K. (1973), pp. 286-324.

Merton, R. K. (1968). The Matthew Effect in Science. *Science* 159, No. 3810 (5 January 1968), pp. 56-63. Reprint in: Merton, R. K. (1973), pp. 439-459.

Merton, R. K. (1969). Behavior Patterns of Scientists. Kappa-Sigma Xi address before the Association for the Advancement of Science in December 1968, published in *American Scientist* 58 (Spring 1969), pp. 1-23. Reprint in: Merton, R. K. (1973), pp. 326-342.

Merton, R. K. [ed.] (1973). *The Sociology of Science: theoretical and empirical investigations*. Chicago, IL: University of Chicago Press.

Miller, J. G. (1978). *Living Systems*. New York, NY: McGraw Hill.

Miller, J. G. and Miller, J. L. (1990). Introduction: the Nature of Living Systems. *Behavioral Science,* Vol. 35, 157-196.

Miller, J. G. and Miller, J. L. (1992). Greater Than the Sum of Its Parts I. Subsystems which process both matter-energy and information. *Behavioral Science,* Vol. 37, 1-38.

Minsky, M. (1985). *The Society of Mind*. New York, NY: Simon & Schuster.

Mullins, N. C. (1968). Social Origins of an Invisible College: the Phage group. Paper presented to the American Sociological Association, Boston, August 1968.

Mynatt, E. D., Adler, A., Ito, M. and O'Day, V. L. (1997). Design for Network Communities. Proceedings of CHI 1997, ACM Press, pp. 210-217.

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

Neuwirth, C. M., Kaufer, D. S., Chandhok, R. and Morris, J. H. (1994). Computer Support for Distributed Collaborative Writing: defining parameters of interaction. *Proceedings of CSCW 94,* ACM Press, pp. 145-152.

Newberry, M. (1995). *Katipo: a Web Lurker*. Victoria University of Wellington, New Zealand. URL: http://www.vuw.ac.nz/~newbery/Katipo.html

Newell, A. (1982). The Knowledge Level. *Artificial Intelligence,* Vol. 18, No. 1, pp. 87-127.

Newell, A. and Simon, H. A. (1972). *Human Problem Solving*. Englewood Cliffs, NJ: Prentice-Hall.

Norman, D. A. (1988). *The Design of Everyday Things*. Originally published as: *The Psychology of Everyday Things*. Readings, MA: Addison-Wesley.

Norman, D. A. (1991). Cognitive Artifacts. Carroll, J. M., [ed.] *Designing Interaction: Psychology at the Human-Computer Interface*. pp. 17-38. Cambridge, UK: Cambridge University Press.

Norman, D. A. (1993). *Things That Make Us Smart: defending human attributes in the age of the machine*. Readings, MA: Addison-Wesley.

Nunamaker, J. F., Dennis, A. R., Valacich, J. S., Vogel, D. R. and George, J. F. (1991). Electronic Meeting Systems to Support Group Work. *Communications of ACM*, July, Vol. 34, No. 7, pp. 40-61. Preprinted in: Baecker, R. M. [ed.] (1993), pp. 718-739.

Olson, G. M. and Atkins, D. E. (1990). Supporting Collaboration With Advanced Multimedia Electronic Mail: the NSF EXPRES project. Galegher, J., Kraut, R. E. and Egido, C., [ed.], pp. 429-451.

Perrochon, L. (1995). A Reference Architecture for Multi-Author World-Wide Web Servers. *Proceedings of COOCS 95*, ACM Press, pp. 197-205.

Phares, E. J. (1988). *Introduction to Personality*, 2/E. Glenview, IL: Scott, Foresman & Co.

Plotkin, H. C. [ed.] (1988). *The Role of Behavior in Evolution*. Cambridge, MA: MIT Press.

Polley, R. B. (1989). Operationalizing Lewinian Field Theory. In: Lawler, E. J. and Markovsky, B. [eds.] *Advances in group processes: a research annual*, Vol. 6. Greenwich, CN: JAI Press.

Popper, K. (1972). *Objective Knowledge: an evolutionary approach*, (revised 1979). Oxford, UK: Oxford University Press.

Rafaeli, S. (1988). Interactivity: From new media to communication. In: Hawkins, R. P., Wiemann, J. M. and Pingree, S. [eds.] (1988). *Sage Annual Review of Communication Research: advancing communication science*, Vol. 16. pp. 110-134, Beverly Hills, CA: Sage.

Rafaeli, S. and Sudweeks, F. (1997.) Interactivity on the net. In: Sudweeks, F. McLaughlin M. L. and Rafaeli, S., [eds.] (1997).

Rafaeli, S. McLaughlin, M. L. and Sudweeks, F. (1997). Introduction Chapter. In: Sudweeks, F. McLaughlin M. L. and Rafaeli. S., [eds.] (1997).

Reid, E. M. (1991). *Electropolis: communication and community on Internet relay chat*. Honour thesis. Dept. of History, University of Melbourne, Australia.

Resnick, P. (1997). Filtering Information on the Internet. *Scientific American*, March, Vol. 276 No. 3, pp. 62-64.

Resnick, P. and Varian, H. R. (1997). Recommender Systems. *Communications of the ACM*, March, Vol. 40, No. 3, pp. 56-58.

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, and Riedl, J. (1994). GroupLens: an open architecture for collaborative filtering of netnews. *Proceedings of CSCW 94*, ACM Press, pp. 175-186.

Rice, J., Farquhar, A., Piernot, P. and Gruber, T. (1996). Using the Web Instead of a Window System. *Proceedings of CHI '96*. New York: ACM, pp. 103-117.

Ritzer, G. (1992). *Sociological Theory*, 3/E. New York, NY: McGraw-Hill.

Roberts, L. G. (1988). The ARPANET and Computer Networks. In: Goldberg, A. [ed.] (1988), pp. 143-167.

Rogers, E. M. (1995). Diffusion of Innovations, 4/E. New York, NY: The Free Press.

Schatz, B. R. (1991). Building an Electronic Scientific Community. *Proceedings of 24th Annual Hawaii Int. Conference on System Science 1991*, Vol. 8, No. 3, pp. 87-107. Reprinted in: Baecker, R. M. [ed.] (1993), pp. 550-560.

Schopler, J. (1965). Social Power. In: Berkowitz, L. [ed.] *Advances in Experimental Social Psychology*, Vol. 2. New York, NY: Academic Press.

Schwartz, M. F. and Wood, D. C. M. (1993). Discovering Shared Interests Using Graph Analysis. *Communications of the ACM*, August, Vol. 36, No. 8, pp. 78-89.

Scott, W. R. (1992). *Organizations: rational, natural, and open systems,* 3/E. Englewood Cliffs, NJ: Prentice Hall.

Shardanand, U. and Maes, P. (1995) Social Information Filtering: Algorithms for Automating “Word of Mouth”. *Proceedings of CHI ‘95*, ACM Press, pp. 210-217.

Shaver, K. G. (1987). *Principles of Social Psychology*, 3/E. Hillsdale, NJ: Lawrence Erlbaum Associates.

Shaw, M. L. G. and Gaines, B. R. (1996). WebGrid: Knowledge Elicitation and Modeling on the Web. *Proceedings of the WebNet96 Conference*, San Francisco, CA., Oct. 15-19, Association of Advancement of Computing in Education (AACE). pp. 425-432.

Smith, J. B. (1994). *Collective Intelligence in Computer-Based* Collaboration. Hillsdale, NJ: Lawrence Erlbaum Associates.

Sowa, J. F. (1984). *Conceptual Structures: Information Processing in Mind and Machine.* Reading, MA: Addison-Wesley

Specter (1995). *WebWatch*. Specter Communications. URL: http://www.specter.com/

Sproull, L. and Kiesler, S. (1991). *Connections: new ways of working in the networked organization*. Cambridge, MA: MIT Press.

Star, S. L. and Puhleder, K. (1994). Steps Toward an Ecology of Infrastructure: complex problems in design and access for large-scale collaborative systems. *Proceedings of CSCW '94*, ACM Press, pp. 253-263.

Stefik, M. [ed.] (1996).*Internet Dreams: archetypes, myths, and metaphors*. Cambridge, MA: MIT Press.

Sterling, B. (1993). Short History of the Internet. *The Magazine of Fantasy and Science Fiction*, February, 1993. Cornwall, CT: F&SF.

Sudweeks, F. McLaughlin, M. L. and Rafaeli, S., [eds.] (1997). *Network and Netplay: virtual groups on the Internet*. Menlo Park, CA: AAAI/MIT Press.

Suen, H. K. and Ary, D. (1989). *Analyzing Quantitative Behavioral Observation Data*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Tracy, L. (1989). *The Living Organization: systems of behavior*. New York, NY: Praeger.

Walster, E. H., Walster, G. W. and Berscheid, E. (1978). *Equity: theory and research.* Boston, MA: Allyn & Bacon.

Wasserman, S. and Faust, K. (1994). Social Network Analysis: methods and applications. Cambridge, UK: Cambridge University Press.

Wellman, B. (1998). Structural Analysis: from method and metaphor to theory and substance. In: Wellman, B. and Berkowitz, S. D., [eds.] (1988), pp. 19-61.

Wellman, B. and Berkowitz, S. D., [eds.] (1988). *Social Structures: a network approach*. Cambridge, UK: Cambridge University Press.

Wheeler, W. M. (1911). The Ant-Colony as an Organism. *Journal of Morphology*, vol. 22, no. 2, pp. 307-325.

White, D. R. and McCann, H. G. (1988). Cites and Fights: material entailment analysis of the eighteen-century chemical revolution. In: Wellman, B. and Berkowitz, S. D., [eds.] (1988), pp. 380-400.

Wiener, N. (1948). *Cybernetics*, (republished in 1961). Cambridge, MA: MIT Press.

Williams, D. (1996). A Brief History of the Internet and Related Networks. http://wwwbs.wlihe.ac.uk/~williams/history/brief_his.html

Williams, G. C. (1966). *Adaptation and Natural Selection: a critique of some current evolutionary thought*, (republished in 1996). Princeton, NJ: Princeton University Press.

Wilson, E. O. (1971). *The Insect Societies*. Cambridge, MA: Harvard University Press.

Wilson, E. O. (1975). *Sociobiology: the new synthesis*. Cambridge, MA: Harvard University Press.

Wilson, E. O. (1992). *The Diversity of Life*. New York, NY: W. W. Norton & Co.

Zuckerman, H. and Merton, R. K. (1971). Patterns of Evaluation in Science: institutionalization, structure and functions of the referee system. *Minerva* 9, No. 1 (January 1971), pp. 66-100. Reprint in: Merton, R. K. (1973), pp. 460-496.

Zuckerman, H. and Merton, R. K. (1972). Age, Aging and Age Structure in Science. In: Riley, M. W., Johnson, M. and Foner, A. [eds.] *A Sociology of Age Stratification, Vol. 3 of Aging and Society*. New York, NY: Sage. Reprint in: Merton, R. K. (1973), pp. 497-559.