UNIVERSITY OF SUSSEX

SPRU - SCIENCE AND TECHNOLOGY POLICY RESEARCH

The Structure of Electronic Scientific Communication

Electronic Networks, Research Collaboration and the Discovery of Digital Knowledge Bases

Alexandre Paulo Fernandes Varela Simões Caldas

A Thesis Submitted in Partial Fulfilment of the Requirements for the Degree of Doctor of Philosophy in Science and Technology Policy Studies.

Brighton, 2004

I hereby declare that this thesis has not been submitted, either in the same or different form, to this or any other University for a degree.

Signature: ______________________________

To my lovely wife Susana and my son Henrique

To my mother and to the memory of my grand-mother

ACKNOWLEDGEMENTS

A work of this nature would not have been possible without the contributions of many other persons and entities beyond the personal efforts of the author who was responsible for creating a report on the results of the research.

I am forever grateful for the love and care of my wife Susana who always encouraged and supported my work and at the most painful times had a word of companionship and motivation. She accepted the responsibility of raising our son Henrique, even at a period of extreme fatigue and stress. She is of infinite value.

I am also grateful to my supervisors, Ed and Ben, who, in adddition to the best possible supervision, provided me with friendship, support and motivation, as well as the most outstanding guidance in becoming a free thinker and an autonomous researcher.

I acknowledge the financial and institutional support of the Foundation for Science and Technology - Ministry of Science and Technology in Portugal, which granted me a research fellowship for the whole period of the investigation (Grant reference PRAXIS XXI/BD/21517/1999), supporting a full time research period of three years in the UK. I also acknowledge the institutional support of SPRU as well as the coordinating service of the ELSnet network - the European Network of Excellence in Speech and Language, for their endorsement of the electronic survey.

I am indebted to all my family for their support, words of motivation and encouragement. I also thank my good friends and colleagues at the Ministry of Science, Paula Almeida, Luís Filipe, Miguel Oliveira, Joana Montalvão and Fátima Biscaia for their continuing companionship, exciting discussions about our work in science and technology policy and personal encouragement. My good friends Adão Carvalho and Vicente Ortiz, at SPRU, provided me with friendship and enlightened experience about the challenges of research. A special thanks to Susana Jorge who with dedication supported the T Media project on Internet consultancy in Portugal while I was away. I am grateful to Cynthia Little who edited and proofread the final manuscript. I am also grateful to many other friends, colleagues and students who have contributed to my personal motivation and curiosity about the advance of knowledge and E-Science.

UNIVERSITY OF SUSSEX

Alexandre Caldas Doctor of Philosophy

The Structure of Electronic Scientific Communication

Electronic Networks, Research Collaboration and the Discovery of Digital Knowledge Bases

ABSTRACT

This investigation focuses on analysing the influence of advanced Information and Communication Technologies (ICTs) and particularly Internet technologies on the process of scientific communication and collaboration. New technologies are likely to create environments where many of the traditional norms of science no longer apply, suggesting that in these new digital environments, scientific communication is more open, pluralistic and interdisciplinary. However, more extensive empirical evidence is needed on the functioning of the phenomenon widely labelled as the “Global Research Village”.

Notwithstanding the extensive research on scientific communication, collaboration and digital infrastructures for science, less attention has been given to the changing patterns of communication among researchers within electronic environments. What are the essential characteristics of the “structure” of electronic scientific communication? Are these electronic patterns of interaction similar to the traditional (non-electronic) structures of scientific communication and collaboration? This research seeks a better understanding of this transformation and discusses the following hypotheses. The first is that the diverse set of Internet technologies is used in multiple modalities across a wide spectrum of communications and collaborations, from the more formal to the more informal transactions, as well as from individual to highly collaborative research activities. This contradicts traditional models of electronic scientific communication that approach the use of information technologies essentially as tools and resources appended to a linear research process. Secondly, traditional patterns of scientific communication and collaboration are being reproduced within electronic environments. This is linked to the hypothesis that the Internet does not radically transform hierarchies established by traditional means, but contradicts the alternative hypothesis that the major impact of Internet technologies is the creation of new and revolutionary forms of research organisation. Thirdly, we investigate the hypothesis that it is possible to identify “digital knowledge bases” - large-scale and distributed electronic knowledge structures - constituting not merely repositories of knowledge and information resources, but reproducing the interactions among research actors and representing electronic patterns of collaboration and communication.

The above research questions and hypotheses are empirically examined by analysing, particularly in Europe, patterns of communication and collaboration in the field of computational speech and language. This investigation develops new methods and indicators for exploring patterns of collaboration in electronic environments. The research combines traditional methods for understanding scientific communication - bibliometrics, network analyses and surveys - with innovative methods for understanding electronic communication - cybermetric techniques for mapping electronic networks and analysing Newsgroup interactive environments. The empirical evidence supports the hypotheses put forward in this thesis and refutes alternative hypotheses derived from the literature.

TABLE OF CONTENTS

INDEX TO FIGURES AND TABLES

FIGURES

TABLES

1. Introduction and Synthesis

First thoughts and motivation

The *Information Revolution* and the Internet have the potential to contribute significantly to the formation and development of "*knowledge societies*" - societies in which the production, distribution and use of knowledge assume a central and systemic role. The extent of this contribution is strongly dependent both on the implementation of technological conditions and on the organisational (re)structuring of society activities and processes - and consequently of the organisation of e-science, electronic science.

Science and its actors - researchers, research groups and research institutions - have had, and continue to have, an important role, whether by contributing directly to the development of these information and communication technologies, or by being early adopters of these innovative technologies, or participating in new forms of organisation and new techno-organisational systems. This investigation focuses on the analysis of these structural transformations and the comparison of traditional patterns of research collaboration with potentially new patterns of electronic collaboration. "Structure" emerges from enduring patterns of relationships among entities, whether organisms, species, and the natural world, or researchers, research groups and the science system. Thus, electronic "structures" emerge as the result of patterns of electronic interactions. In this sense, the adoption of a structural perspective in the analysis of change of research collaboration and communication, and resulting modes of organisation, into electronic infrastructures is deliberate.

For this structural perspective I am indebted to the work of two outstanding twentieth century scientists: Thomas Kuhn, and his landmark studies on the structure of scientific development and Stephen Jay Gould, biologist and paleontologist, and his remarkable studies on the structure of evolutionary theory. The common element in these analyses of evolution in biological and scientific systems is the importance of "structure" - patterns of interaction among entities.

While we are probably still at an early stage in the information revolution in science, this research analyses these structural transformations and seeks to advance our understanding of the "structure" of electronic scientific communication.

1.1 Electronic Information Infrastructures for Science

The availability of advanced Information and Communication Technologies (ICT) has the potential to have a significant influence upon the science system. In fact, scientific progress is influenced by the structure of scientific communities, and the potential of electronic information infrastructures to influence this structure is large. These "electronic information infrastructures" comprise the whole set of technologies and systems which support the processing, storage and interchange of data, information and knowledge. Our particular focus is on communication and collaboration technologies and the Internet. We investigate how patterns of scientific communication and research collaboration are transformed as they are transferred to electronic environments. This transformation comes from the dynamic interaction between the technological infrastructure on the one hand and the socio-organisational science infrastructure on the other.

It has been a central science policy goal to maximise the benefits of the use of digital infrastructures, electronic networks and applications for scientific research. Most recently, the capability of electronic networks for supporting the work and collaboration of social networks (hence "e-science networks") is providing the motivation for a closer investigation of the social organisation of e-science. In this regard, this thesis contributes to an advance in knowledge by approaching these electronic networks as effective social networks of communication and collaboration. To make this contribution, this research is aimed at achieving a better understanding of the electronic patterns of interaction that are emerging. The thesis will provide answers to the following research questions: Are traditional patterns of research collaboration and scientific communication being reproduced in electronic environments? If so, what is the extent of this reproduction?

Theoretical and empirical studies concerning the process of scientific communication, the organisation of science and scientific communities, the quantification and measurement of scientific performance, the comparative performance of diverse forms of organisation, the networking of research resources and the incentives and motives for collaboration in science, provide a very good framework for understanding the socio-organisational science system (for a literature review, see sections 2.1.2 through 2.1.4 in the conceptual framework set out in chapter 2). Several complementary aspects of the electronic technological infrastructure have already been identified (see e.g. OECD, 1999, 2000; PITAC, 2001 and for a more thorough discussion see sections 2.1.5 and 2.1.6 of the conceptual framework chapter in this thesis). Among these, the following are worth noting here: the existence of advanced physical electronic networking infrastructures and services for the scientific and technical communities; the establishment of high-performance computing and communication research environments to support large-scale and distributed computing; the standardisation and application of GRID[1] architectures and services which are likely to provide hardware, software and middleware for research computing and collaboration; the provision of advanced electronic networking for the remote control of scientific instruments and access to and dissemination of scientific data; the wide dissemination and use of digital libraries and the transformation of systems of electronic publishing in science; and, finally, the structural analysis of changes in communication among scientists and patterns of use of these new information technologies and electronic networks.

Notwithstanding the above elements of an e-science "system", significant differences exist between Europe and the United States (and even within countries across Europe) with regard to the establishment of these digital infrastructures for science. Research into electronic networking and patterns of ICT use among research communities in Europe is urgently needed. Following from the identification of gaps in the literature and in order to answer the research questions posed in this thesis, it is clear that some research issues require more thorough explanation. Three particular characteristics of e-science features require closer investigation.

First, we need more evidence on the regularities, intensity and patterns of use of ICTs and Internet technologies for research, communication and collaboration. These electronic collaboration activities, supported by a diversified set of technologies, are likely to produce a wide spectrum of formal and informal interactions, in activities ranging from the more personal and individualistic to the more collaborative. Second, we need a better understanding of the transfer of traditional research structures of communication and collaboration into electronic environments and of whether traditional patterns of collaboration are being reproduced in electronic infrastructures, or whether these important transformations accompany this transfer process. Several dimensions are worthy of detailed analysis: the *connectivity* of various forms of organisation and interaction, the *specialisation* of collaborating groups (and the division of labour), the centrality and prestige of different collaborators (and *hierarchy* in science), the *scope of interaction* (e.g. inter-sectoral, international, etc.), the *universality* (public dissemination of information, accessibility to

[1] GRID computing can be defined as applying resources from many computers in a network - at the same time - to a single problem; usually a problem that requires a large number of processing cycles or access to large amounts of data. Source: IBM GRID computing - available at the following URL: http://www.ibm.com/grid last accessed 2003-02-12.

confidential data, extent of interaction within the community, etc.), as well as the *evolution* of these electronic interactions (formation, development and disruption). Finally, we need to understand the extent of public dissemination of information within electronic environments, as well as the patterns of electronic interaction and collaboration revealed by these large-scale and distributed electronic knowledge structures. Are we able to identify "digital knowledge bases"? If so, do they reveal patterns of interaction similar to non-electronic forms of communication and collaboration?

This thesis aims to contribute to a better understanding of these transformations by analysing empirical evidence of e-science in the field of computational speech and language, with a particular focus on European research. The empirical evidence produced within the lines of investigation suggested above, leads to a discussion of an "electronic research collaboration" model (proposed in section 2.3 of the conceptual framework chapter and thoroughly discussed in the conclusions to this thesis - see chapter 6).

The following sections of this chapter introduce the topics for investigation. In the first place, we analyse the socio-organisation infrastructure of science. Section 1.2 discusses the process of scientific communication and patterns of formal scientific communication. Section 1.3 examines the networking of research resources in non-electronic settings and discusses non-electronic forms of research collaboration. We also analyse how these socio-organisational science systems might be influenced by the use of technological infrastructures. Section 1.4 interprets the change to electronic environments in structural terms as a change in patterns of connectivity among entities (researchers, research groups and research institutions). Finally, sections 1.5, 1.6 and 1.7 focus on the details of technological infrastructure (understanding these electronic networks as supporting science networks). Section 1.5 discusses patterns of use of ICT and Internet technologies for research communication and collaboration. Section 1.6 assesses the reproduction of research collaboration structures in electronic environments. And section 1.7 raises the issue of the identification of large-scale digital knowledge bases within electronic environments.

1.2 Scientific Communication and Patterns of Research Collaboration

Scientific communication is the essence of science (Garvey, 1979). The process of research is widely believed to be incomplete without the publication of its results. Moreover, scientific communication serves the additional purposes of validating knowledge claims and archiving the knowledge accumulated within the edifice of science. The emergence of electronic infrastructures is likely to change traditional modes of scientific communication and collaboration and we intend to provide a model of the emergence of these new connections and new interdisciplinary communities.

Notwithstanding the usual emphasis attributed to the formal system of communication and publication in scientific journals, the process of scientific communication covers a wide spectrum of activities, ranging from the formal to the informal, i.e. face-to-face communication, activities within collaborative projects, exchange of ideas through meetings, correspondence or pre-prints within closely knit, densely connected "invisible colleges", knowledge inter-change during conferences, seminars and symposia, the interchange of personnel and several other forms of research interaction.

The structure of the formal system of scientific communication is better understood than the more "invisible" and informal communication and knowledge exchange occurring between researchers, research groups, institutions or whole research communities. Bibliometric and scientometric methods exist for mapping the structure of "formal scientific communication". Particular techniques for producing the quantitative measures of science are applicable for mapping "research collaboration" - such as co-authorship, even if this is considered to be only a partial indicator of collaborative activity. On the other hand, methods for mapping informal scientific communication are not so common.

As our focus is on the organisation of scientific communities, it is of special concern here to understand the "structures of collaboration" emerging from formal publications. Only a long-term analysis of formal communication is likely to reveal the inherent collaboration structures in a field of research. However, there are some problems with bibliometric and scientometric studies, such as under-counting or under-identification of collaborative activity. Moreover, as scientific communication practices and norms vary significantly across disciplines and fields of research, particular problems are likely to occur when assessing collaborative structures in inter-disciplinary, application-oriented fields. Several methods are proposed to overcome some of these problems. First, the identification of research groups/institutions, which, through an extensive long-term analysis of bibliometric data, are likely to reveal more intensive collaborative activity. Secondly, analysis of the structural impact on these networks of the collaboration of particular "leading" researchers, whose seminal works have contributed in a revolutionary way to the growth and development of fields of research. Thirdly, connectivity and centrality measures, which are likely to vary significantly across different collaborative groups, might identify alternative modes of organisation within these scientific communities as well as revealing hierarchies of organisation within those communities. Finally, studying the evolution of these structures of collaboration is likely to reveal important characteristics of the evolution of the research communities and the field of research (e.g. it may allow identification of research groups that are able to conduct seminal work and stimulate the emergence of new research areas or the fundamental restructuring of existing areas - such as in the case of speech recognition, see results chapter 4.1).

Knowledge of the structures of collaboration, as given by the formal system of communication, is important for understanding the transfer of traditional features of communication and collaboration into electronic environments and the accompanying transformation of collaborative activities. The forms of such communication are evolving due to the existence of electronic infrastructure, and it is important to understand whether these new forms change the structure of scientific communities - i.e. their hierarchy, the ways in which workgroups are connected, the division of labour, etc.

A review of the research literature on the topic of quantitative measures of scientific communication and collaboration structures is provided in sections 2.1.2 and 2.1.3. More detailed analysis of the bibliometric and co-authorship methods is provided in sections 3.1 and 3.2 of the methodology chapter. Section 4.1 discusses in detail the results of the empirical examination of bibliometric data for the research area of computational speech and language - the area under analysis in this research.

Analysis of the formal system of scientific communication reveals the "tip of the iceberg" in the spectrum of communication and collaboration in science. An in-depth study of research collaboration activities and structures of research collaboration is also needed. The next section introduces the mapping of collaborative structures as a way of identifying patterns of connectivity among research collaborators, as well as potential hierarchies within scientific communities. The empirical evidence comes from an analysis of research networks funded by the European Union and data on collaboration activities of individual researchers obtained from a survey of researchers participating in a European research network - the ELSnet network (see methodology chapter, section 3.1).

1.3 Research Networks as "Social Structures" of Scientific Collaboration and Communication

The collaborative nature of science dates back to the beginnings of the modern era, but much more recently there has been an upsurge of interest in science policy initiatives for networking research resources. In the last two decades of the twentieth century, research networks emerged as empirical "forms" of collaboration that integrate various research resources into wide networks of collaboration. These "networking" initiatives have the potential to strengthen the connectivity of research collaboration structures as well as to "institutionalise" more informal collaboration practices. All this is likely to contribute to increased communication within the boundaries of the collaboration network, as well as between external entities, linked by one or more associative relationships.

Research collaboration assumes various forms and occurs at several levels of aggregation (Katz and Martin, 1997), knowledge about which is crucial for understanding research organisation and research performance. There are different, but inter-connected, levels of collaboration such as inter-personal research collaboration, inter-group collaboration, inter-departmental, inter-institutional and inter-national collaboration. Collaborative research also covers a wide spectrum of activities, from the informal domain of short-term personal exchange of preliminary ideas on a research project, to institutionalised large-scale big science projects, involving inter-disciplinary research resources, spanning international boundaries, and undertaken over extended periods of time.

Extensive research has highlighted the motives and causal factors for research collaboration efforts (e.g. Katz and Martin, 1997; David, Foray and Steinmueller, 1997; OECD, 1999; ESF, 2000; EC, 2000 and Widhalm *et al.*, 2001). Some of these studies focus on explaining the motives and incentives for collaboration - in other words, "*why* collaborate?", whereas others focus on a better understanding of the structures and patterns emerging from collaboration - "*How* is collaboration carried out?" Explanatory factors for collaboration include the growing complexity of research problems, which require more research resources, greater division of labour, wider inter-disciplinary knowledge and more extensive long-term commitments; reduced funding for fundamental research, driving the need for pooling and networking research resources; and the improvement in transport and communication capabilities that facilitate remote collaboration between geographically dispersed research groups. Studies of the forms of collaboration include the quantification of collaborative activities - bibliometric and scientometric analyses; the formation and development of collaborative activities - evolution of collaboration; and the detailed analysis of research funding and its impact on the structure and networking of researchers and research groups; or even sociological research into the norms and values underlying scientific collaboration and the priority-based reward system in research leading to hierarchies in scientific organisation. An extensive discussion of both thematic areas is provided in sections 2.1.1 through to 2.1.4 of the conceptual framework chapter.

Some specific characteristics of these collaboration structures are worth detailed empirical analysis at the level of the research group/institution as well as at the level of individual researcher collaboration. First, what are the medium to long-term patterns of connectivity emerging from stable collaboration practices (*e.g.* in collaboratively funded research projects)? It is likely that connectivity patterns and the centrality and prestige of research entities will vary significantly among members of these networks. A thorough analysis of these effects is provided in the chapters dealing with the long-term network analysis of research groups participating in european funded speech and language research contracts (see section 4.2). These network dynamics might affect the whole structure of these collaboration systems, as suggested by the new economics of science (Dasgupta and David, 1987, 1994). Secondly, what effects do individual researcher collaboration practices have on the whole structure of the research community? Collaboration structures emerging from collaboration between individuals are likely to be related to inter-institutional practices and these forms of organisation have implications for the performance of the science system. Several other factors might affect this relationship, such as inter-sectoral and international patterns of collaboration, how collaboration patterns vary according to the "seniority" of the researchers involved, and the inter-disciplinary background of researchers and research groups. Chapter 4.2 discusses in detail the empirical evidence on patterns of collaboration among researchers and research institutions in the European computational speech and language research area (based on survey data and the partnerships resulting from collaborative funded research projects).

The analysis focuses on the structural characteristics of networks of collaboration. Particularly relevant is the connectivity and centrality of individual scientists and research institutions, the subgroup structure of these networks, and the evolution or dynamics of these structures over time. The empirical evidence characterises the traditional structures of collaboration and scientific communication, but is also likely to contribute to a better understanding of the adoption of electronic networks for research work. In fact, the typical structures of research communication and collaboration (developed into features such as the reward structure in science and the Matthew effect[2] or Lotka's Law,[3] the progressive specialisation of science activities and "division of labour", the inter-sectoral and international nature of collaborative activities, or the subgroup and clustered-structure of scientific organisation) are likely to be reproduced in environments of electronic collaboration.

The integration of these structures of collaboration and communication with the potential offered by ICTs and electronic networks will determine the extent of success of "e-science" and the "global research village".

1.4 The Structure of Electronic Scientific Communication

In this section we specifically discuss the structural perspective adopted in this thesis, by analysing the interaction between the socio-organisational features of science (introduced in the previous two sections) and electronic information infrastructures (to be introduced in the following three sections). This thesis treats electronic networks as infrastructures that effectively support the work and collaboration of scientific networks. In this sense, these electronic networks are more like social networks - groups of researchers, research groups and research institutions, and their communication and collaboration links. This structural perspective attributes great importance to the nature of the collaborative links (their extent, connectivity, evolution, specialisation, etc.) as the research issue that is worth detailed analysis. This analysis requires information about each researcher (actor) in terms of collaborative behaviour and use of ICT for collaboration, as well as information on the collaborative activity of the community taken as a whole techno-social network (large-scale non-electronic and electronic collaborative activity).

---

[2] The *Matthew effect* as originally formulated by Merton (1968b) had a strict meaning - within the formal publication system in science and the traditional priority-based reward system of science - of progressive accumulation in prestige resulting from the previous publication track record of individual researchers. Later formulations by Merton's followers widened the scope of meaning to refer more generally to the cumulative nature of the reward system in science.

[3] Lotka (1934), while investigating the productivity of individual researchers and the variability of their research productivity within scientific communities, postulated an *inverse power-law* relationship, still widely used in other fields, between the number of researchers in a community and the productivity (as measured by the number of articles in scientific journals) of that community. This leads to a simpler formulation that a large part of scientific production is attributable to the work of just a few highly prolific authors, whereas the majority of researchers make a relatively minor contribution to total output.

A better understanding of the process of electronic scientific communication and the emergence of new structures (patterns of electronic interaction) require both theoretical and empirical knowledge on the personal patterns of use of information and communication technologies (as evidenced by individual scientists) on the one hand, and on the other hand, on collective communication and collaboration activities within electronic networks as evidenced by research groups, research institutions and entire research communities. As science is widely considered to be a collaborative endeavour, the most dramatic effects of ICTs on science are likely to occur within the more collaborative electronic activities. Structures or patterns of research collaboration and communication are thus likely to be manifested within these research electronic environments. Social "structure" emerges from enduring patterns of collaboration and communication among entities - whether individual researchers, research groups or research institutions.

When one adopts a structural perspective in the analysis of these electronic infrastructures, specific properties of these social networked systems gain particular significance. First, particularly important is the *inner* structure of these networks (e.g. size, number and diversity of members, number and diversity of cohesive subgroups whose interactions are tightly bounded and densely connected). Secondly, the connectivity patterns of individual nodes, subgroups or clusters, or even of the whole network. merit thorough analysis. Particularly important is the variability of these connectivity indicators, as this might reveal transformations in non-electronic forms of research collaboration. Thirdly, the significance of the centrality and location of particular entities within these electronic networks, and the consequent effect of this centrality on their importance in structuring the network should be examined, as this might reveal hierarchies in the organisation of the research community. Fourthly, the evolution and dynamics of these electronic networks in the short, medium and longer term, and particularly their structural dynamics in terms of specialisation of components and subgroups within the whole network systems merit close investigation. A better understanding of these electronic patterns of collaboration and their characteristics is likely to reveal features of scientific community organisation such as the division of labour and specialisation of research groups, intersectoral and international collaboration, hierarchies in the community and variability in scientific reputation and prestige.

In line with this structural perspective, the crucial research questions then are what characteristics are intrinsic to the 'structure(s)' of electronic scientific communication? Are the patterns of electronic collaboration similar to those embodied in traditional (non-electronic) scientific communication and collaboration? If so, to what extent is this reproduction occurring? There are a number of key issues to investigate. The first is concerned with the technological characteristics of these electronic environments and the extent of their usage by scientists. This interaction between technological conditions and scientific work and collaboration, encompasses a wide variety of activities from formal publication to informal transactions, and individual practice to highly intensive collaboration. The second relates to the extent of reproduction of traditional features and structures of non-electronic research communication and collaboration in electronic environments. The extent of replication is strongly dependent on the degree to which these electronic networks are able to support science (social) networks. The third issue concerns the emergence of new knowledge structures within these large-scale and distributed electronic networks, resulting from practices of public dissemination of information and electronic knowledge interactions.

The conceptual framework underlying this investigation is discussed in chapter 2. In particular, section 2.2 analyses in detail the research question and theoretical hypotheses to be empirically validated. The methodology chapter (chapter 3) describes the methods and techniques used for collecting and analysing the empirical evidence.

It is widely believed that information and communication technologies have a significant impact on the functioning of scientific communities and on their composition and organisation. However, the extent of this effect is strongly dependent on the specific characteristics of the process of non-electronic scientific communication as well as the traditional research collaboration structures of the field of research under investigation. While sections 1.2 and 1.3 briefly introduced these two socio-organisational dimensions, sections 1.5, 1.6 and 1.7 introduce respectively the patterns of use of ICTs and Internet technologies for research work and collaboration; ways in which the replication of traditional forms of organisation and hierarchies in electronic environments can be identified; and the criteria on which the emergence of innovative forms of electronic knowledge interchange can be ascertained.

1.5 Information and Communication Technologies and Changing Patterns of Communication among Scientists

A better understanding of the nature of traditional structures of communication and research collaboration is a fundamental first step for assessing the extent to which these patterns of interaction are being transformed into electronic forms of networking. Another complementary dimension for gaining insight into the transition to electronic environments is knowledge about the pattern of usability of these technological infrastructures. These usability patterns influence the extent to which collaborative activities are indeed occurring within electronic environments, and the extent to which new distributed knowledge structures are emerging within these electronic networks as a result of the *digitalisation* of traditional collaborative activities and structures.

Chapter 5 examines patterns of use of information technologies by researchers for the purposes of research work, communication and collaboration (section 5.1), the use of electronic interactive environments (i.e Newsgroups) for informal scientific and technical collaboration (section 5.2), and the extent to which Digital Knowledge Bases - highly distributed, large-scale electronic knowledge structures - are indeed emerging within electronic networks (section 5.3). The methodology chapter (Chapter 3) discusses the data sources for the empirical analyses conducted in chapter 5, namely electronic survey data from the ELSnet network (the research network of European speech and language research), archived messages of speech newsgroups over the period (1992 - 2000), and World Wide Web (WWW) digital information for the ELSnet electronic network. Sections 1.5, 1.6 and 1.7 briefly introduce each of these research topics.

Empirical evidence of the use of ICTs and electronic networks (e.g. the Internet) for research work, communication and collaboration is fundamental for gaining insights into the structural transformation of these processes, both technological and organisational. This is particularly true in the case of European research communities, where the provision of electronic information infrastructures for science is lagging behind those in other world regions (e.g. the United States, Canada and Japan). In this regard, an important part of the empirical evidence in this thesis was obtained from an electronic survey of scientists carrying out research on computational speech and language across Europe.

The results of the web-based survey give important insights into the diversity of the technologies subsumed under the label "ICTs", the wide spectrum of applicability of diverse technologies, as well as the *multi-modality* with which technologies are used for different dimensions of formal/informal communication and for inter-personal/collective collaborative activities. The term *multi-modality* in this thesis is used to represent the wide spectrum of applicability with which different technologies are being used for different dimensions of research work, communication and collaboration. The same technology (e.g.

e-mail or newsgroups) is likely to be used in a wide spectrum of communication (formal or informal) and collaboration (individual or collaborative) activities, but the extension of this usability is also likely to vary significantly across technologies. The approach and results of this survey contrasts with traditional models and approaches to electronic scientific communication that have focused either on analysis of the impact of a single technology (e.g. e-mail, ftp, etc.) (see e.g. Bar-Ilan, 1997; Ginsparg, 1997 and Walsh and Bayma, 1996) or on the extension of the traditional model of scientific communication with different technologies (see e.g. Crawford *et al.*, 1996; Howells, 1995 and Kling and McKim, 1997), but not on the transformation of the research process resulting from the use of these technologies, or the intertwining of the technological and organisational networks.

Following this alternative approach, in which electronic networks support science (social) networks, an important topic for investigation is the extent to which these technologies are being used to support "extended research groups" - remotely dispersed forms of organisation for research activities. In addition, these electronic networks are also likely to be used to support local research activities.

Also, one would expect a certain amount of variability in the intensity and diversity of use of technologies among researchers within any particular community. What is still an unknown is whether there are regularities in the usage patterns of those technologies, whether there are systemic differences in their use in different aspects of research work, including communicative and collaborative activities. The latter issue is referred to in this thesis as "multi-modality". Which technologies are being used for co-authoring publications, for working on research projects, for exchanging research results with close collaborators or with the scientific community at large, for accessing information on complementary specialties, research funding or new research contacts? The spectrum of research activities supported by different technologies is likely to emerge from the patterns of use related to each of these activities.

In addition, it is of particular value from the individual perspectives of researchers, to understand the extent to which electronic networks are used as repositories of knowledge and as environments for knowledge interchange. Defining this is likely to yield empirical evidence of the diversity of information being disseminated and exchanged within electronic environments. To what extent are institutional resources, such as research group contacts, research project documentation, pre-prints and refereed publications, being disseminated openly and publicly within these infrastructures? It is particularly important to understand how these electronic resources reveal connectivity with other collaborating research groups (e.g. electronic co-authored publications, collaborative projects, direct links to close collaborating research groups, etc.). This electronic connectivity might reinforce the reproduction of traditional patterns of collaboration in these electronic environments.

Finally, how are and for what purpose are more advanced research computational and communication resources being used? These advanced electronic services include remote access to large digital databases, fully functional electronic publication systems, remote access to computational servers for sophisticated numerical analysis, or even electronic services supporting on-line collaboration. The intensity of use of more advanced resources provides evidence for the usefulness of more advanced information infrastructural initiatives in certain fields of science.

The personal (and collective) patterns of usage of information and communication technologies and electronic networks are also likely to be reinforced by analysis of long-term archival records of electronic collaborative systems. The next section introduces the exploration of electronic discussion fora as one such e-collaboration techno-organisational system and electronic environment for knowledge interchange. Section 1.7 describes how electronic networks function as repositories of knowledge resources and infrastructures supporting collaborative interactions.

1.6 Electronic "Invisible Colleges" and the Process of "Division of Labour" in Science

While evidence about personal use of electronic networks reveals regularities in usage patterns, detailed examination of more interactive electronic environments is likely to reveal structures of collaboration. Electronic specialist scientific and technical discussion fora are likely to constitute ideal environments for informal scientific communication and collaboration.

Structures of traditional (non-electronic) communication might or might not be reproduced within these environments. The structural evolution of electronic systems has the potential to reveal important characteristics such as progressive specialisation and "division of labour" and inter-sectoral and international collaboration. The electronic networks might also be structured in ways that are similar to traditional collaboration groups. Researchers and research groups might also reveal different levels of activity and prestige within electronic environments. And, finally, the content of discussions might not be properly considered as "scientific knowledge", but only inter-change of technical or personal non-validated information. Thus, a typology of the content being exchanged is needed.

There is not much empirical evidence on the detailed mapping of the structural evolution of electronic systems of interaction (e.g. newsgroups). Are traditional patterns of diffusion, such as the logistic growth curve, suitable for modelling the evolution of these systems? More importantly, to what extent do these systems vary in the long-term in terms of size (number of discussions, number of contributors, number of specialised fora), complexity of structure and diversity of content? Can we identify highly differentiated groups of contributors in terms of activity and prestige? Is the relative importance of discussions evenly distributed or significantly skewed?

A second focus is on the degree to which these electronic systems support inter-sectoral and international collaboration. To a large extent this would be revealed by the number and diversity of contributors to the interactive discussions.

A third topic for analysis is the degree to which these electronic systems reveal the formation and development of structures of collaboration. Extensive network analysis of the collaboration networks resulting from long-term electronic connectivity is likely to reveal the subgroup structure of these systems. Can we identify densely connected, highly interacting and tightly bounded cohesive subgroups? If so, is there any pattern of specialisation among those subgroups? In analysing the structural evolution of newsgroup systems as a whole, can we validate the "division of labour" phenomenon?

Finally, detailed content analysis of newsgroup messages is likely to reveal a typology of the content being exchanged within these electronic environments. Although these electronic environments lack a formal system of refereeing and knowledge validation, the degree of technical and scientific information being communicated should be a reasonable indicator of their ability to be a medium for scientific communication and collaboration.

If electronic networks and interactive collaborative environments support the traditional structures of informal scientific communication and collaboration, a more detailed analysis of the electronic connectivity of researchers and research institutions may reveal the emergence of large-scale distributed knowledge structures within these electronic networks. This analysis is introduced in the next section.

1.7 Electronic Connectivity and the Discovery of Digital Knowledge Bases

Electronic networks might support the communication and collaboration activities of large social networks (in our case, research networks in various forms). If we can empirically validate the previous hypotheses about the reproduction of traditional structures of scientific communication and collaboration in digital environments, this will constitute a necessary but not sufficient condition for these knowledge collaboration structures to be able to disseminate the results of research activities across electronic space. These electronic knowledge structures consist of personal interactions among researchers, collaborations among research groups and research institutions and the activities of large research networks.

The second and complementary condition of stability required in collaborations among research entities (at various levels - e.g. researcher, research group, research institution) and in the complex and highly negotiated processes of the production, distribution and use of scientific knowledge, should also be manifest in long-term electronic collaboration activities.

The existence of large-scale, highly distributed knowledge structures - digital knowledge bases - would require the validation of the above two conditions, as well as open and public dissemination of knowledge resources across these wide area electronic networks. The organisation of electronic web space may be consistent with the organisation of traditional research collaboration. These similarities may be assessed in terms of the connectivity patterns of closely collaborating institutions, the long-term collaboration activities of these same institutions and the enduring practice of electronic dissemination of their research results. If so, it may be possible to discover and characterise such electronic knowledge structures.

It is worth noting from the outset that a complete evaluation of these electronic digital knowledge bases is almost impossible given the variety of ways that information may diffuse through the social network and may influence behaviour. In these circumstances, what can be done is to assess whether the structures involve relatively random interconnections or, alternatively, whether they reproduce the structures of collaboration and interaction that can be traced using other means (bibliometrics, research funded networks, collaborative projects, etc.). If the latter applies, there is a basis for concluding the evaluation positively, altough not definitively - alternative explanations like the possibility that such electronic structures are simply the by-product of generous research funding require more thorough analysis. This thesis examines empirical evidence on the electronic connectivity of the research institutions belonging to the ELSnet research network. Particular emphasis is given to testing the following hypotheses. First, that the pattern of electronic connectivity among these institutions is structurally similar to the identified non-electronic collaboration structures within the ELSnet network. Secondly, that there is an element of self-organisation to these electronic networks, and a detectable sugbgroup structure with some research institutions having higher levels of activity and better connectivity, which is being more centrally located within the whole network. Finally, the dissemination of knowledge resources in these public and open electronic networks by the best connected and most active research institutions is taken to be a demonstration of the existence of electronic knowledge structures.

To sum up, this investigation collected and analysed empirical evidence on patterns of use by scientists of ICTs and particularly Internet technologies for research work, communication and collaboration. This evidence is used to evaluate the extent to which electronic interactive environments are reproducing traditional forms of scientific communication and collaboration as well as providing the basis for identifying digital knowledge bases - electronically distributed knowledge structures. The next section provides an overview of the organisation of the dissertation.

1.8 Organisation of the Dissertation

This section briefly summarises the structure of the dissertation. Chapter 1 has outlined the research topics covered in the investigation. Each research topic has been briefly summarised including the fundamental sub-topics, hypotheses and results.

Chapter 2 provides a detailed analysis of the conceptual framework which sets the boundaries for the current research. After an extensive analysis of the literature on scientific communication, research collaboration and information infrastructures for science (section 2.1), the research questions and theoretical hypotheses under investigation are examined in detail in section 2.2. The structure of electronic scientific communication and the extent of its similarity to traditional structures of scientific communication and collaboration, as well as the associated hypotheses, are discussed in this section. A new conceptual model for electronic research collaboration, integrating the hypotheses to be examined, is introduced in section 2.3.

The methodology chapter (Chapter 3) describes the combination of methods and techniques used for collecting and analysing the empirical evidence on electronic scientific communication. The following methods are examined: bibliometric analysis of patterns of formal scientific communication; network analysis of patterns of institutional research collaboration in collaborative projects and patterns of electronic connectivity within newsgroups and among research institutions within electronic web space; study of the characteristics of the electronic survey carried out to collect evidence on researchers' interdisciplinary backgrounds, individual researchers' collaboration practices and scientists' usage patterns of ICT. In addition, the use of the innovative cybermetric methods developed during the investigation is explained, on the one hand, to analyse knowledge interchange through newsgroup interactions and, on the other hand, to understand the electronic connectivity of research institutions on the World Wide Web and the extent of knowledge dissemination.

Chapter 4 discusses the results of the examinations of patterns of non-electronic scientific communication and patterns of non-electronic research collaboration. These examinations are conducted in two complementary sections, 4.1. and 4.2. Topics covered in section 4.1 include the bibliometric delimitation of the boundaries of inter-disciplinary fields of research, the structural impact of revolutionary contributions, the mapping of structural evolution of collaboration networks and the identification of collaborative groups based on co-authorship networks. Section 4.2 discusses the evolution of patterns of institutional collaboration networks, the different patterns of connectivity of research institutions and patterns of connectivity within inter-personal research collaboration.

Chapter 5 presents a discussion and the results of the examination of the structure of electronic scientific communication and electronic research collaboration. Patterns of information and communication, particularly Internet technology, usage by researchers are discussed in section 5.1. In this section, the following topics are considered: electronic networks as a medium for "extended research networks"; multi-modality in the use of different technologies; electronic networks as repositories of knowledge and environments for knowledge exchange; and patterns of use of the more advanced electronic networks and services. Section 5.2 discusses in depth the quantitative and qualitative results of the analysis of interactive and collaborative newsgroups, particularly the evolution of these electronic systems, and potential processes of specialisation and "division of labour", quantitative indicators of activity and prestige, structures of collaboration emerging from informal communication within these environments (resembling "invisible colleges") and an embryonic typology for the information exchanged within electronic discussion fora. The discussion of electronic connectivity between research institutions and the identification within these large-scale and sparse electronic networks of digital knowledge bases are provided in section 5.3. In this section we analyse methods for mapping electronic networks on a large-scale and distributed networks and patterns of electronic connectivity within electronic networks, as well as the identification of differentiated centrality and prestige within these electronic environments. Lastly, a conceptual model for identifying digital knowledge bases is presented.

Finally, Chapter 6 summarises the main conclusions of the investigation and discusses lines of further research. This chapter includes: a synthesis of the empirical findings; discussion of the fundamental theoretical conclusions; re-examination of the conceptual model of electronic scientific communication and research collaboration; discussion of the limitations of the investigation; and references to complementary lines of research.

The references used and the Annexes to the dissertation are included in the final part of the thesis.

2 Conceptual Framework

2.1 Literature Review

The aim of this investigaton is to understand how the emergence of electronic scientific communication is influencing the structure of collaboration and networking among researchers. In particular, we focus on the changing patterns of communication and collaboration among scientists and the potential replication of hierarchical structures that may occur as a result of the use of advanced ICTs and particularly Internet technologies. Understanding these changes requires a thorough analysis of two complementary and interacting and influential dimensions: on the one hand, the technological capabilities of electronic networking; and, on the other hand, the socio-organisational processes of research interaction and collaboration. A systemic analysis of these changes requires an approach which considers electronic networks as social networks (of researchers, research groups and research institutions).

The fundamental research question under examination is whether or not essential characteristics of the structure of traditional (non-electronic) communication are reproduced in electronic environments of scientific communication and collaboration. A comprehensive answer to this question requires assessment of three complementary dimensions of these environments. First, there is a need for better understanding of the processes of scientific communication - taking these processes as covering a broad spectrum from formal to informal activities - that can be gained from the quantification and characterisation of the research collaboration structures emerging from these processes (sections 2.1.2 and 2.1.3 review the literature on this dimension). Secondly, the organisation of research communities and the social structures characterising research collaboration should be thoroughly analysed. This analysis must take into account a wide range of the "forms" of collaborative activities, including science policy efforts aimed at supporting the networking of research resources (the literature review relating to this dimension is contained in section 2.1.4). Thirdly, digital infrastructures for science must also be taken into account. These infrastructures include a wide range of activities from electronic networking infrastructures, high-performance computing and communication infrastructures, remote access to scientific instrumentation and data sharing, electronic publishing and digital libraries, "collaboratories", and changing patterns of communication among scientists. Section 2.1.5 briefly summarises the research within the generic theme of "digital infrastructures for science". Section 2.1.6 focuses on a more detailed analysis of research on the specific topic of this investigation - how patterns of communication among scientists change with the availability of electronic networks and large-scale, distributed computational resources, such as the Internet.

Analysis of the literature reveals some gaps in our knowledge which this investigation will try to fill. Following from the literature, and within the specific framework of electronic networks supporting research collaboration (hence science networks), there is a lack of understanding and detailed empirical examination of the particular properties of these electronic interactions. These properties include: firstly, how research actors (researchers, research groups and research institutions) use these electronic networks for supporting research collaboration across a range of formal and informal and individual or collaborative activities, given the wide diversity of available technologies; secondly, to what extent traditional forms of research collaboration and patterns of research communication are being reproduced in electronic environments; thirdly, to what extent long-term electronic interaction and public dissemination of information and knowledge, allow the identification of digital knowledge bases - that is, large-scale and distributed knowledge structures on the Internet.

A detailed discussion of the research question and specific hypotheses under investigation follows the literature review and synthesises the contributions to follow. Section 2.3 discusses the formulation of a new conceptual model for analysing electronic research collaboration. This model functions as an integrative conceptual framework for the empirical analyses described in chapters 4 and 5 of the thesis.

The discussion starts with an analysis of scientific communication and proceeds to the analysis of research collaboration and digital infrastructures for scientific communication and collaboration and particularly electronic networks supporting science networks. After the literature review, the research question and hypotheses are discussed and the chapter concludes by proposing a conceptual model for electronic research collaboration, leading the way to the examination of the methodology.

2.1.1 Initial Remarks on Science Studies, Collaboration in Science and the Use of Information and Communication Technologies

From the debates of the 1960s and 1970s concerning social studies of science, and the more focused and pragmatic endeavours in the 1980s of researchers engaged in the sociology of scientific knowledge (SSK), a large body of knowledge has been created on how scientific knowledge is produced, distributed and used. For the present research purposes, these social science studies could be categorised in terms of two major, and essentially independent, strains of investigation: on the one hand, the practitioners of the history and philosophy of science, whose primary concerns were focused on the content and epistemological characteristics of science and its outcomes (see, e.g. Popper, 1959 and 1963; Quine, 1962; Kuhn, 1962; Lakatos, 1970); and on the other hand, the sociologists of science, in the functionalist tradition of Rober K. Merton, who focused on the social context, embedding individual scientist's attitudes and behaviour, namely the normative structure of science and its priority reward system (see, e.g., Zuckerman and Merton, 1971; Crane, 1965; Cole and Cole, 1967 and 1973; Hagstrom, 1965; Whitley, 1984; Ziman, 1984). From this historical perspective, the SSK practitioners attempted to bridge between the two separate sets of studies, focusing on the social and performance inter-relationships of the scientific system. Within the tradition of the sociology of scientific knowledge (SSK), recent studies on the structure and behaviour of research networks have attempted to shed light on how scientists collaborate and how science is conceived within its broader social context (for a critical overview, see Knorr-Cetina, 1995; Callon, 1995 and David, Foray and Steinmueller, 1997). Drawing on a wide range of disciplines, from the history of science and sociology of science, to empirical studies of science as practice, these studies have produced a rich menu of conceptual and methodological practices, as well as providing the seedbed for new approaches to improve the understanding of scientific culture and practice. A demonstration of the variety (and, to some extent, disunity) of the various approaches is exemplified in the sequence of articles from the more realist, empirical approach (Collins and Yearley, 1992), to the reflexivist perspective (Woolgar, 1992), and proponents of the actor/network model (Callon and Latour, 1992).

A diversity of approaches characterises these science studies and establishes a general framework for analysing research collaboration. Some of these studies have focused on the nature of innovation systems embedding research collaboration processes (along the lines of Kline and Rosenberg, 1986; Freeman, 1991; Lundvall, 1993; and Gibbons *et al.*, 1994). Others have focused on examination of the forms of organisation emerging from research collaboration and the particular incentives and motives leading to collaboration, in line with the "new economics of science" (e.g. Dasgupta and David, 1987; David, 1995; and David, Foray and Steinmueller, 1997). Another strand of science studies focuses on the quantitative analysis of collaboration and the statistical and empirical examination of research collaboration structures and performance (following bibliometric and scientometric studies, Price, 1963, 1965 and 1987; Katz and Martin, 1997; and Katz and Hicks, 1997). Yet another set of science studies focuses on the particular characteristics of research collaboration, such as the transdisciplinary nature of reseach collaboration or the networking of research resources via new electronic communication and collaboration infrastructures (e.g. Callon and Latour, 1992; Knorr-Cetina, 1995 and 1996; Galison, 1997; Roosendaal, 1997; and Meadows, 1998). This section reviews some of this literature, particularly focusing on "why does reseach collaboration occur" and then analysing the particular features of organisation of research collaboration and "how research collaboration is shaped and structured".

Starting in the late 1980s and extending through the 1990s, a significant number of science studies specifically focused on the resource allocation factors characterising the collaboration practice of researchers within specific research networks. These networks are considered as forms of organisation of the wider scientific community in a certain research field, and these studies have offered a new framework for contributions in the "new economics of science" tradition (see e.g. Dasgupta and David, 1987 and 1994; David, 1995; David and Flemming, 1996; and Baños *et al.*, 1999). The point of departure in these studies that differs from previous approaches is twofold. First, it involves the identification of units of analysis at a higher level than the individual scientist, in order to study how science is organised and how well these social organisations perform (considering that most of them never become institutionalised) in the pursuit of the advancement of knowledge. Second, it offers a renewed focus on the science side of the science and technology equation, with novel theoretical and empirical investigations into the social organisation of scientific communities and their performance.[4] Several studies have been conducted analysing research networks as collaborations of education and research institution within the European Framework Programmes (Geuna, 1997; Garcia Fontes and Geuna, 1997; Removille and Clarysse, 1999; Luukkonen and Hälikkä, 2000). More analyses of collaboration practices between universities and industry in the more global tradition of

[4] About the unbalanced efforts in the investigation of science communities organisation and performance, and resource allocation efficiency considerations, as opposed to technological studies on the same conditions, see e.g. Dasgupta and David (1987, 1988), David (1994) and Diamond (1996).

innovation studies have been carried out (see, for example, Kline and Rosenberg, 1986; and Freeman, 1991).

This current research is directly focused on the nature of the communication links established within these research networks, the structure of these interactions and the influence of ICTs and Internet technologies in changing the communication process. The empirical investigation in this thesis of the speech and language research community aims to contribute to a better understanding of scientific organisation and performance.[5]

One set of studies that is more related to the research described in this thesis focuses on the collaboration practices of scientists and the statistical study of performance indicators, and also involves the mapping of informal communication links within the wider scientific community (or "invisible college") and the establishment of local social networks or reference groups. Following the work of Price (1963, 1965, 1986) on the statistical analysis of publication performance within research communities, more recent advances have come from the conceptual and methodological analysis of collaboration as reflected in scientific publication (see e.g. Katz and Martin, 1997[6]). The continuing work of the so-called "translation school" based upon the initial work of Callon (1989, 1991, 1992) has attempted to link these performance or output analyses (using publication and patent indicators) with the network analysis of informal communication processes within communities of practitioners, in order to gain a better understanding of the influence of various forms of research collaboration organisation on research performance. The integrated use of these statistical and econometric procedures as well as network analysis methodologies is of paramount importance for mapping the dynamics of these flexible networks of scientific practitioners.

Yet another set of recent theoretical and empirical investigations adopts a more "transactionist approach" and focuses on knowledge accessibility and the communication process within local networks of researchers, and on the implications of the normative structures and reward systems prevailing in these networks for the performance of the scientific community as a whole (see e.g. Cowan and Foray, 1997; David and Foray, 1994, 1995, and 1996; David and Flemming, 1996; David, 1995, 1997, and 1998; Arora, David

[5] For an analytical and historical treatment of networks among research scientists and engineers, see e.g. Antonelli (1994), Dasgupta and David (1994), David (1995, 1997), and David and Flemming (1996).

[6] This research provides us with a useful taxonomy for defining "research collaboration" focusing essentially on collaboration in scientific publication at various levels: individual, research group, departmental, institutional, sectoral and national, as well as highlighting the partiality of using bibliometric indicators such as co-authorship to measure research collaboration.

and Gambardella, 1998). Paul David's recent work on communication norms and the structured characteristics of research networks, and the implications for the epistemic performance of invisible colleges, provides an approximate framework for analysing in greater detail the scientific communication process within advanced information and communication technologies environments (David, 1998).

Two more considerations frame the theoretical and empirical investigations which are relevant to gaining an understanding of the scientific communication process: the composition of the network, in other words the multi-actor nature of research networks and the potential combination of different normative and reward structures influencing the behaviour of research practitioners (for example public researchers belonging to the education and public laboratories systems working in combination with private business researchers), and how this may affect the overall performance of the community;[7] and a better understanding of the transdisciplinary nature of the research process and how this is reflected in the structure of the network in the first place, and the behaviour and performance of its practitioners as an outcome of particular relevance to those concerned with improving the efficiency of the scientific communication process.

On this latter topic, several "science as practice" studies (see e.g. Callon and Latour, 1981 and 1992; and Knorr-Cetina, 1992, 1995 and 1996) conducted during the early 1980s and 1990s have set science practitioners' collaborations within a boundary, neither following the traditional rigid disciplinary-based approaches, nor conforming to completely relativist frameworks. Several accounts, ranging from more theoretical conceptions to specific field work and empirical studies, have tested these approaches. According to one school of thought, the production of new knowledge, no longer occurs only inside disciplinary boundaries. It also occurs in the interstices between established disciplines, through the cross-fertilisation between disciplinary areas, and through the diffusion of instruments and procedures which affect the practice of research in other remote areas (Gibbons, 1994). Another approach is based on the emergence of trans-disciplinary concepts such as elastic

[7] Macro-level analysis of the social negotiations taking place within a multi-actor knowledge production system is naturally left out of the current investigation. Nevertheless, at micro-level, it is extremely important to understand the incentive structures in local research networks and the scientific community as a whole. For a "controversial" macro-level account, see e.g., Gibbons *et al.* (1994). The "Mode 2" model of knowledge production is referred to in Gibbons *et al.*, who describes the emergence of a socially distributed knowledge production system, within which Knowledge is produced under an aspect of continuous negotiation (Gibbons *et al.*, 1994). If the university settings are likely to lose their dominance over knowledge production, what new systems are likely to emerge? Is there likely to be a system built in which the institutions of higher education, the universities in particular, will comprise only, but perhaps only a small part, of the knowledge production sector? (Gibbons *et al.*, 1994). If so, what negotiation strategies and what structural changes are likely to occur in the scientific communication process?

boundary objects (Star and Griesemeyer, 1989) which promote our understanding of translation efforts in the management of collective work across scientific disciplines. Likewise, Fujimura (1992) has used the concept of standardised packages, to investigate cancer research as a global collaborative enterprise involving basic research (in fields such as genetics, virology, molecular biology and neuroscience), government laboratories and institutes and several other interests. The concept of "standardised packages" unifies the common practice of cancer research, linking the various actors while leaving their essential properties unaffected, but allowing scientific communication to be carried on (Fujimura, 1992). Several other accounts, from a more experimental and "laboratory studies" perspective, could be presented in the same vein, from the laboratory studies of Knorr-Cetina (1992, 1995, 1996), or Hacking, (1992, 1996) or even experimental practice (Gooding, 1992). Another interesting study is that of Galison (1997) in his account of twentieth century physics, within which theory, experiment and instruments are brought together in a "trading zone", where theorists, experimenters, and engineers, in a locally coordinated setting, develop a "trading language" (pidgin or creole) to bring about the scientific development of physics knowledge. All these arguments for a more flexible approach to fields of inquiry and application have deep consequences for the scientific communication process, both directly as regards the language mechanisms and procedures, and in terms of the multi-actor and negotiation techniques used, and thus the diverse patterns of collaboration.

The above literature review summarises a diversity of studies focused on answering the "why research collaboration" question, while a subset of these studies assumes collaborative processes as a given and specifically focus on explaining particular forms of organisation, patterns and structures of collaboration - in other words, "how collaboration operates".

Given the general framework of theoretical and empirical investigations conducted in the analysis of research collaboration networks and scientific communication, as well as the multi-actor nature and transdisciplinarity characteristics of this scientific collaboration process, two aspects focus this research endeavour. The first is the imbalance regarding the understanding of the formal system of scientific communication relative to informal communication and collaboration activities. The second is the relative absence of studies specifically focusing on the impact of electronic environments on the functioning of scientific communities.

In the first place, far more is known about the formal communication subsystem of scientific practice, as it is materialised through the normal publication activity of the research community, as well as through participation in institutional bodies and institutionalised meetings, conferences and other events. Knowledge has also been accumulated through investigations into the influence of ICTs in general and of electronic communications in particular on the formal scientific communication system (for a recent account, see e.g. Roosendaal, 1996, 1997; Geurts and Roosendaal, 1998; Shaw and Moore, 1996; Meadows, 1998; Kircz, 1998). In the Mertonian tradition of the sociology of science (Merton, 1973), the claim for priority in science has been recognised as part of a collegiate reputational reward system to support the global advancement of knowledge, and the success of science as an autonomous social activity. Scientific publication (and public disclosure of new knowledge) has been used as a self-regulating mechanism since the beginnings of modern science in seventeenth century England, and the appearance of the first "scientific journal". The emergence of electronic environments as settings for scientific communication is likely to produce radical changes in the overall system of normative values and other contexts of science practice. On the other hand, far less is known about the informal system of communication - in the form of personal contacts, correspondence, the interchange of messages between research practitioners, and so on. The significance and importance of tacit knowledge (Polanyi, 1963) as a local and contextualised body of skills, practices and procedures available, but not fully explorable, within scientific exchange processes, has also been emphasised in several studies.[8] The validity of these arguments has had a deep influence on the communication system within electronic environments. On the one hand, if information and communication technologies do facilitate the codification of knowledge (Steinmueller, 1998) in its various dimensions as an aid to individual memory, an aid to group memory and an aid to social memory, then electronic environments are likely to facilitate the overall process of scientific communication. On the other hand, recognising the importance of informal processes of communication (skill transfer, personal contacts, conference participation) is likely to demand more flexible electronic environments with higher levels of distributed capacity and non-deterministic behaviour.

Secondly, only a few theoretical and empirical investigations have been concerned[9] with, and much less is known about, the influence of information and communication

[8] For a comprehensive synthesis of the research network literature, see David, Foray and Steinmueller (1999).
[9] In this regard, the small number of such studies in Europe is striking, when comparing to parallel efforts in the US - see e.g., National Research Council (1994), National Research Council (1995).

technologies and electronic environments on the functioning of scientific communities and, specifically, the expected changes in the scientific communication process and the global performance of these communities.[10]

A classification of these ICT developments and their benefits in scientific communication appears to provide a convenient point of departure for the current investigation. Moreover, the research process and, in terms of this particular investigation, the nature of the scientific communication process, are likely to change given the networking, simulation and animation potentialities of advanced ICTs. Given the possibilities opened up by the astonishing development of networking technologies over the last decade, it seems reasonable to expect changes in the possibilities for scientific communication and consequently the structure and dynamics of research collaborations and research networks. This research aims to contribute to advancing our knowledge about these aspects.

The following two sections review in detail the literature on scientific communication and collaboration (section 2.1.2) as well as on quantitative analysis - biblometric studies - of the formal system of communication in science (section 2.1.3) in providing evidence on collaboration patterns. Section 2.1.4 extends the discussion of communication and collaboration to specific forms of organisation of scientific communities and particular forms of collaboration involved in networking research resources. Finally, sections 2.1.5 and 2.1.6 complement the previous discussions of the socio-organisational features of the science system with a focus on the technological dimension. The literature on digital infrastructures for science is reviewed and examined as a means to understand the electronic networking of research resources.

[10] The literature reveals the remarkable coincidence of a reinvigorated interest in studies concerning research networks and the application of the network metaphor to a better understanding of scientific communities, with the recent development of information and communication technologies, see e.g. Dasgupta and David (1994). This research is an attempt to provide a better understanding of how ICT developments are used in such research networks, and of the optimisation of this inter-relationship.

2.1.2 Scientific Communication and Collaboration and Scientific Communities

'James D. Watson and Francis Crick discussed on the BBC how they made their discovery of the molecular structure of DNA, a discovery which led them, along with Maurice Wilkins, to receive Nobel Prizes in 1962. In the course of their discussion they recalled the time when Maurice Wilkins had in his possession an X-ray photograph of DNA in the B form, which he was trying to keep Linus Pauling, Watson and Crick from seeing until he had a chance to exhaust his own study of the photograph. "But", Crick said, "he had to eventually make public the photograph **because communication is the essence of science**"'. (emphasis added)

(Garvey, 1979, page xi)

The above quotation, beginning the discussion in this literature review section, emphasises the importance of formal scientific communication to the whole edifice of science. As Francis Crick reported at the time of his revolutionary discovery of the structure of DNA, made public in a journal article co-authored with James D. Watson and published in *Nature* in 1953, 'communication is the essence of science'. In addition, the quotation above underlines the strong inter-relatedness between the scientific research process, its activities and social organisation collaboration structures, and the end-products of those activities - scientific information in the form of published journal articles. With regard to this viewpoint, Bodin, when discussing the essence of scientific information, considers scientific information in the form of publication as being inseparably linked to the social activity of "scientific research" (Bodin, 1989). From a similar perspective, Derek de Solla Price makes the statement that 'the process of scientific discovery is incomplete without publication' (Price, 1980: 221). Moreover, he connects this publication rationale with the 300 year-old process of claiming priority for a discovery. As such, scientific communication in the form of publication is not only a medium for communicating scientific discoveries, but also a historically proven and established normative feature of scientific practice.

Recognising the historical importance of scientific communication in the form of publication for both the archiving of "reliable scientific knowledge" and the social organisation of scientific activity constitutes a significant first step for comprehending the scientific communication process and the need for collaboration in science. An additional step is to recognise that beneath the term "scientific communication" or "communication in science" lies a diverse set of activities and processes. For distinguishing among this wide spectrum of activities, Garvey and Griffith provided in the 1970s a very robust, and hitherto unchallenged, conceptual model. In this traditional model of scientific communication, the formal activities of publication represent but one extreme of a wide spectrum of other communication activities, having at the other extreme the informal domain of science (Garvey and Griffith, 1972). In a brief account of studies of scientific communication Griffith (1989) summarises the contribution of this model in three complementary dimensions. First, it provides a model for the disciplinary processing of information. Second, it serves as a model distinguishing between the formal and informal realms of dissemination of scientific information in terms of both function and structure. And, thirdly, it constitutes a model which characterises the role of the productive scientist as driving and controlling scientific information-exchange.

In fact, the importance of informal communication practices and loosely connected networks of communication among researchers is now commonly accepted. Research in this vein goes back to the late 1960s, and particularly to Menzel's insistence on bringing the informal domain into the mainstream of studies of scientific communication (Menzel, 1968). He briefly summarises the regularity of some of these important informal communication activities, and the position of certain scientists in acting as "the scientific troubadour".[11] Such scientists frequently act as carriers of information from place to place, recipients of correspondence, hosts to visiting scientists and as visitors to other institutions. Menzel also includes other functions such as editors of journals, members of grant-application review committees, or organisers of and participants in "summer laboratories or courses". He emphasises the frequency with which informal networking activities take place, for instance at summer laboratories, in the corridors during scientific meetings, and during and after colloquia and conferences.

Other studies have emphasised the collective and collaborative character of such loose networks in contributing to the advance of research (e.g. Price and Beaver, 1968; Crane, 1972; and Cole and Zuckerman, 1975). These "invisible colleges" - tightly bounded but loosely connected networks of "peers" - are considered to be of fundamental importance to the practice of research. Pioneering examples of these studies are Crane's (1972) analysis of rural sociologists and mathematicians, the study of the information exchange group on oxidative phosphorylation and terminal electron transport carried out by de Solla Price and Beaver (1968), and Cole and Zuckerman's (1975) study of the invisible college of sociologists of science). The common element in all these studies is that they all understand scientific communication as both a social process and a structure of organisation of scientific communities.

[11] When revising this reference, the personality of Sir Robert Moray appeared to me as a "scientific troubadour", when in the period 1650 to 1662, he "bridged" communication links among various other members of an "Invisible College" who had their meetings in Gresham College, the place of the origins of the Royal Society of London. Part of this historical account can be found in Lomas, Robert (2002).

A recent survey of studies of scientific communication (Lievrouw, 1998) highlights the wide spectrum of approaches that must be taken into account when studying communication in science. Lievrouw's typology is based upon the degree of "commoditisation" of scientific information and communication. At one end of the spectrum, we have the *Artefact studies* (e.g. bibliometric and scientometric studies), in which scientific information is taken as an "objective commodity, whose value is independent of its use" and therefore may be rigorously measured and quantified. At the other end of the spectrum, in *Laboratory studies*, scientific information is taken as "a social construction of scientists", and only ethno-methodological studies are seen as able to collect useful data on the real practice of science and on where and how science is done. In the middle part of this spectrum lie *User studies* and *Network studies*. The former class of investigations considers "scientific information as a commodity whose value depends on the practical needs of the user". User surveys on how users of scientific information seek and collect this kind of information and indeed *use* it for the practice of science is a typical methodology for this kind of research. On the other hand, *Network studies* usually embed the scientist and practitioner in a web of associations (network) where scientific information is taken as 'a social link, whose value is determined by its utility in the coherence of social networks' (Lievrouw, 1998:12). Network analysis is the typical methodology for this kind of investigation. As we move from the artefact studies to constructivist laboratory studies, scientific communication is taken to a higher degree as a dynamic process, rather than as a static property of scientific activity, and we reach a more practical definition of this process as a process of negotiation and interpretation of meanings by the individuals involved. However, this move also increases the problems of "measuring this process" and of rigorously quantifying its characteristics.

Two essential features of the process of scientific communication are highlighted in the above literature. First, the wide spectrum of formal and informal activities going under the label of "scientific communication". Secondly, the collaborative nature of this process, and its being a social activity involving communities of researchers and research institutions. The next section briefly reviews some techniques now widely used for measuring the structures of collaboration emerging from the communicative process of science and identifying patterns of collaboration characteristic of the development of research areas.

2.1.3 Bibliometric and Scientometric Indicators: Measuring the Formal System of Scientific Communication

The above review of the literature sets the general framework for discussing the use of bibliometric and scientometric methods for understanding the structure of research collaboration. The following paragraphs discuss in more detail the main arguments put forward in terms of the advantages and limits of these quantitative studies.

Biliometrics or Scientometrics now constitutes a field of research in itself. The pioneering studies on bibliometrics of Derek de Solla Price and by E. Garfield in the 1960s (see e.g Price and Beaver, 1968; and Garfield, 1979) spurred an interest in the quantitative study of science. In a pioneering study, Garfield, Sher and Torpie analysed the citations to a wide range of research papers on DNA in order to locate the most central authors, and compared their list with Asimov's historical account of the development of the area (Garfield *et al.*, 1964). Since then, the field of bibliometrics has grown substantially with various applications of bibliometric analyses to a diverse set of research areas, and for several quantitative analysis purposes. The field of bibliometrics has its own established and specialised scientific journals (e.g. *Scientometrics* and the *International Journal of Scientometric and Cybermetrics*), as well as its own international meetings and symposia (e.g. International Conference on Scientometrics and Informetrics and the Leiden conference on Bibliometrics and Scientometrics). The Bibliometrics field of research is firmly grounded in the disciplines of information science and the sociology of science. More recently, it has developed strong connections with the emerging areas of "digital libraries" as well as "cybermetrics" research, two areas that are developing as independent fields of interest. Research in digital libraries is focused on analysing *managed* electronic information systems supporting the dissemination of data, information and knowledge. On the other hand, cybermetrics research is about replicating some of the same measurements of intellectual proximity and "relatedness" as bibliometrics using different objects (electronic entities such as web pages, electronic transactions or web information) and different ideas about the nature of "linkages".

The quantitative focus of bibliometric and cybermetric research studies along with the focus on information retrieval and information management techniques, is thus not surprising. It should be stressed that the more recent availability of computational databases of indexes and abstracts of the scientific literature has marked a significant upsurge in the use of these methods. Paramount among such electronic abstracting services, and the most commonly used for bibliometric analyses, is the Institute for Scientific Information (ISI) set of databases. These include the Science Citation Index (SCI), the Social Science Citation Index (SSCI), and other electronic bibliographic databases[12].

A more technical but brief definition of bibliometric and scientometric studies would include in this category studies based on a methodology combining a set of techniques for the exploration of significant portions of archival scientific publications, for the purpose of identifying communication patterns and behaviour, specific features of the organisation of scientific communities and for mapping the long-term development of research fields. As the process of scientific communication assumes a wide variety of forms (see previous section), the set of bibliometric methods included under the general label of "bibliometrics" consists of a diverse set of techniques, such as bibliographic counts, citation analysis, bibliographic coupling, co-citation analysis, co-word analysis and co-authorship analysis. As our main focus here is on research collaboration, the fundamental bibliometric technique of interest is co-authorship, but we briefly summarise the essential characteristics of all these techniques as they reveal how the organisation of scientific communities, the epistemological development of research fields and the evolution of the research fields constitute important factors influencing research collaboration and, thus, electronic research collaboration. The detailed exploration of publication data is assumed to produce a good understanding of how research groups and particular scientists within research groups collaborate with each other. At a more general level, it shows how the whole scientific community in that field of research is organised into subgroups (some more traditionally aligned with the "research programme", others forming more heterodox sub-communities). *Co-authorship* is particularly suitable for these kinds of analyses, but other techniques include *bibliographic counts* of publications and *citation impact* analyses. Next, the analysis of long-term series of bibliometric data is assumed to provide a good understanding of the epistemological development of the field of research. Moreover, if analysed at different points in time, it is assumed to provide a good understanding of the dynamics of the "research frontier" in a particular field and of the active reseach groups contributing to the advance of knowledge on that research frontier. *Citation analysis*, *co-citation analysis* and *co-word analysis* are some of the techniques applicable for mapping the epistemological development of research fields. Finally, detailed analysis of long-term patterns of collaboration and association in formal scientific publication is assumed to

[12] As an indication of the coverage of these electronic bibliographic databases the following numbers are illustrative. The SCI extended bibliographic database covers a set of around 5,300 scientific journals in

provide a good understanding of the evolution of collaboration in fields of research and whole research areas.

Nevertheless, as extensive or even longer than the above is the list of criticisms levelled at the use of bibliometric indicators for mapping fields of research, particularly the use of "conceptual maps". David Edge (1979) provides a general overview and survey of the main critical assumptions of bibliometric methods and their inherent limitations. He characterises these methods as probably allowing the representation of the "tip of the iceberg" of scientific communication. Most of his comments draw attention to the extensiveness of the whole spectrum of communication activities not "measurable" by bibliometric techniques, adding criticisms to the assumed "universalistic agenda" of scientometrics and its lack of grounding in a more sociological approach in which scientific consensus constitutes an open question and is not assumed *per se*. Moreover, sociologists often contest the idea that scientists share a common cognition and "understanding" of their fields of research, as is generally assumed by bibliometric studies. The critique against "universalism" includes not only attacking the assumption that the frontier of science is well-defined but also that all potential players might contribute similarly to the construction of the scientific enterprise. Both premises are also related to a critique of the Mertonian norm of "open science". When sociologists emphasise the importance of "local context" and "proximity" for the actual practice of science (in the vein of laboratory studies, see e.g. Latour and Woolgar, 1982; Latour, 1987; and Knorr-Cetina, 1992, 1995, 1996), they usually support their argument with the complexity and variety of scientific practice. They level the most radical criticism to larger scale studies, attempting to identify systematic patterns and regularities, based upon the production and distribution of artefacts such as journal publications, as envisaged by bibliometric studies (see e.g. Banos-Ruiz *et al.*, 1999; Bordons *et al.,* 1995; and Van Raan, 1998, 1999). On the other hand, these larger scale studies that reveal consistent structures and regularities in the social structures of scientific practice may contradict the "particularist" bias of situated studies. Bibliometric studies certainly do not constitute a comprehensive recording of patterns of communicative behaviour of scientific knowledge, but they usually provide enough evidence of regularities and systematic patterns in the social structure underlying scientific communication and collaboration.

In the next part of this section, we discuss some characteristics of bibliometric techniques, and then compare the benefits and advantages implicitly recognised in the literature, on the

several fields of Science. The SSCU covers 1,700 journals in the Social and Behavioral Sciences.

one hand, and the problems and limitations widely attributed to these quantitative approaches, on the other.
A general assumption made by all bibliometric studies is that, given an extensive[13] data source of publication records in a specific research area, it is possible to measure important characteristics of scientific communication, based on the measurement of certain characteristics of those documents (e.g. authors, author's affiliations and cited references).

Among these characteristics are general patterns in the development of the research area, the association between different documents, the association between different authors, research groups and research institutions, as well as the structure of fields of research and the formation and evolution of those structural patterns. These analyses are conducted at several levels of aggregation and in various dimensions, such as publications, scholarly journals, authors, research groups and research institutions, nations or even whole research areas.

A pure and simple *bibliographic count* of the literature would count the number of papers in a given set of publications (presumably defining a research area), along with other characteristics such as the number of co-authored papers - i.e. the number of papers with more than a single author. Based on these data, one might distinguish certain patterns in the evolution of publication activity over a reasonable time period (for applications, see e.g. Price, 1967).

Citation analysis is focused on analysis of the number of citations attributed to a given publication (how many authors have cited this publication in their bibliographic references list). This is crudely assumed to be an indicator of "impact". The most often cited authors are taken as having a major "influence" on other authors working in the specific field of research in which they publish (see e.g. Wouters, 1999). *Direct citation* analyses determine the set of documents that cite another given document over time. This might also be assumed as a "prestige" indicator, if the analysis covers a reasonable period of time.

Bibliographic coupling attempts to determine when two papers cite the same one or more other papers. Based on this association between documents, some association between authors and research groups is assumed, particularly if some regularity in the pattern of association is confirmed over time.

Co-Citation analyses measure the set of documents that are co-cited. Two documents are co-cited if they appear together in the list of references in a research paper. This is usually taken as a measure of association between pairs of frequently cited documents, and consequently a measure of association among the corresponding authors of those documents. The assumption is "strong" in the sense that it takes the strength of co-citation between documents as a measure of the "conceptual proximity" of two or more authors. From these co-cited networks of association, overall "knowledge-maps" are built up, representing the conceptual maps of a community or field of research. The evolution of these "knowledge-maps" gives an overall perspective on the formation and evolution of the research areas under study (it is outside the scope of this investigation to list extensively applications of co-citation analyses, but for some indication, see Cowan and Jonard, 2000 and Haan, 1997, as well as Baños *et al.*, 1999, refs. 1 and 2).

Co-Word analysis measures the relationship between documents, as given by the similarity in frequency of a set of keywords or key-phrases. This technique attempts to take advantage of some of the benefits of "content-analysis" of documents, in order to complement techniques such as co-citation. Based on expert knowledge about a valid set of keywords, the association between scientific publications is considered to indicate "conceptual proximity" of the authors publishing those scientific publications. In this sense, the assumptions are quite similar to co-citation analyses. Similar techniques to co-word analysis have a remote origin in lexicographical works which explain co-occurrence of knowledge produced several decades ago (e.g. Hornby, 1942), and in the linguistics research of De Saussurre (1949) in describing the correlation between the affinity of language units and its appearance in language. More recent research applications for research evaluation purposes can be found in the work of Callon and colleagues (Callon *et al.*, 1979) and for very recent work along these lines see e.g. van Raan (1998, and 1999) and Noyons *et al.* (1999).

Finally, *Co-Authorship* analyses measure the association between authors who appear in the scientific publications as co-authors of a piece of research. It is reasonable to accept that one obvious form of mutual influence among scientists is active collaboration leading to shared-authorship of research papers. If a paper is co-authored, there is presumably a link between the authors. The frequency of co-authorship gives a measure of the strength of association between the authors. Over a reasonable period of time, clusters of closely associated co-authors are considered to indicate denser and stronger research collaboration

[13] Even the most enthusiastic supporters of bibliometric techniques agree that a bibliometric study should be based on a very extensive set of journal articles, or other publication dataset, as a starting point for the

groups. Applications of co-authorship for mapping collaboration patterns, and particularly in inter-disciplinary areas, are reviewed in later paragraphs of this section, and thoroughly discussed in the data results chapter (section 4.1).

The above summary of bibliometric techniques gives an overview of the use of these methods for constructing maps of science; whether overviews of scientific communication practices (bibliographic counts and citation analyses), "knowledge maps" (bibliographic coupling and particularly co-citation and co-word analyses), or "collaboration-maps" (namely co-authorship analyses). A thorough discussion of the advantages, reliability and limitations of these methods is outside the scope of the present investigation, but some relevant considerations are provided in the following paragraphs.

Several studies offer some positive arguments for the use of bibliometric methods in analysing the quantitative nature of scientific communication. For example, Meadows (1974) considers the advantages of the availability of vast amounts of data, as well as the assessment of these large data sets in "objective, quantitative terms" . While taking a cautious position in the use of co-authorship indicators for measuring research collaboration, Katz and Martin (1997) recognise certain advantages of these methods. They are "invariant and verifiable" as they constitute historical archives of scientific publication. Moreover, they also constitute a relatively inexpensive and practical method for quantifying collaboration. Other cited advantages are that the size of the samples available for analyses are relatively large and, thus, more statistically significant than is possible with other collection methods and, finally, that these methods are "unintrusive and non-reactive" and as such do not affect the course of the investigation process. In several complementary studies van Raan provides strong arguments in favour of the use of bibliometric methods for research evaluation, the study of knowledge diffusion, interdisciplinary relations and research collaboration (van Raan, 1998, 1999; Nyons *et al.*, 1999). These arguments reinforce the importance of the extensiveness of the citation databases (i.e. registration of all citation references to other articles, as well as the detailed coverage of the affiliations of authors of scientific publications).

However, the problems and limitations of the use of bibliometric methods have been widely recognised. A very strong argument against the use of these methods across different fields of research is the inherently very different practices of communication between the sciences, the social sciences and the humanities. Some of the factors

---

various scientometric analyses. See e.g. Van Raan (1998, page 4).

influencing these differences were thoroughly analysed some time ago (particularly in the 1960s and 1970s), but they still remain valid; they include lags in the information flow through the publication process, and differences in the organisation and effectiveness of informal networks, as well as the different modes of transferring information from the informal to the formal domain (see Garvey *et al.*, 1970). Even within these disciplinary categories, for example in the social sciences, significant differences exist in publication and citation practices. Van Raan (1998) points out the radical differences in this respect between the strongly internationally oriented experimental psychologists, and the much more locally oriented sociologists. In the same study, he draws attention to the complexity of bibliometric analysis on more inter-disciplinary and industry-oriented fields of research. In the more application-oriented research fields, there is a less open publication practice.

Another interesting point is made in Whitley's (1970) analysis of the British social science journals. Indeed, the scientific journals in some disciplines might not represent the primary medium of scientific communication, even for formal scientific communication. In his analysis the empirical evidence confirmed the lesser importance of scholarly journals as a medium of communication, compared to books. When assessing research productivity based on abstracting and indexing databases, and thus bibliometric analyses, MacRoberts and MacRoberts (1982) point to at least three major problems. First, these abstracted data sources are almost never complete in their coverage of scientific journals. Secondly, they are seldom complete in covering the whole set of publications within their journal set. Finally, they do not cover extensively other formal publication forms, such as books and symposia proceedings published as books.

When discussing in detail the applications of some techniques, we need to consider an extensive list of criticisms or "problematic issues". We present here just two examples, one dating from the late 1980s and one very much more recent; one concerning co-citation analyses and the other co-authorship analyses. In a thorough analysis of citation practices, MacRoberts and MacRoberts (1986) examined papers to see how much "influence" appears as "references" in bibliographies. Their main argument is based on the incompleteness and imprecision of the references cited in papers. They point to several major drawbacks such as the practice of parsimonious citing, ignorance of the literature, mis-citation, the use of secondary sources, and citation for reasons unrelated to intellectual content. Their qualitative assessment is that the list of cited references in a paper represents only about 30% of effective influence. If this is correct, it would clearly be a problem for the "conceptual maps" derived from co-citation analyses. Another study focuses on assessing the reliability of using co-authorship indicators for measuring research collaboration. Katz and Martin (1997) call attention to the technical problems likely to arise in such procedures. The problems are particularly acute when aggregating the basic units of analysis (papers and co-authors) into higher analytical levels of analysis (such as research collaboration between research groups, institutions, sectors and nations). These problems are due to incompleteness and inaccuracy of the information about the authors' institutional affiliations, but principally concern the validity of taking co-authorship as representing effective collaboration among co-authors. Moreover, they point out that honorary co-authorship, or the fact that researchers effectively collaborating might end-up publishing their research results autonomously, or researchers who have barely contributed to the research might end up on a list of co-authors, might lead to inaccurate interpretations, if co-authorship is used as a measure of collaboration. They suggest that co-authorship should be used as only a partial indicator of research collaboration.

Despite the intense debate on the use of bibliometric analyses, and particularly co-authorship analyses, for assessing research collaboration structures, the fact is that the extensiveness of such studies constitutes obvious empirical evidence of the interest ascribed to them. One argument surrounding their limitations is that these studies play on the availability of data for crafting measurement techniques with a dubious connection to the reality of scientific knowledge or the actual sociological structures that they purport to trace. More realistically, if the identification and characterisation of collaboration structures emerging from the formal system of scientific communication is now widely accepted, there is also agreement on the limited extent to which the informal activities and consensus building behaviour of researchers and research communities might be assessed exlusively using scientometric analyses. In other words, and recapitulating the arguments outlined above , bibliometrics does not claim to be a comprehensive way of recording patterns of behaviour in the communication of scientific knowledge, but that it provides enough evidence to reveal regularities in the social structure of these processes and systematic patterns of collaborative and communicative activities.

This leads to a discussion of the investigations focused on the study of social organisation of science activities and the actual formation and development of research collaboration structures above the level of the individual scientist and encompassing the networking of research actors and research resources. The next section reviews this literature.

2.1.4 The Structure of Scientific Communities and the "Research Network"

Studies in the 1960s and 1970s, in the tradition of Robert K. Merton's sociology of science, sought a better understanding of the organisation of science activities and the diversified set of "social structures" supporting scientific collaboration (see e.g. Griffith, 1969; Mullins, 1972; Crane, 1972; and Ziman, 1984). In this conceptual framework, the activity of the individual scientist (researcher) is embedded in the wider social context, which constitutes the scientific environment and potentially determines its scientific achievements. These investigations underlined the importance, when conducting science studies, of taking into account a wide spectrum of "organisational" forms, from the individual scientist at one extreme to completely institutionalised and formalised collaboration structures, such as scientific societies, consortia and research networks, at the other extreme.

The collaborative nature of science was certainly reinvigorated by the beginnings of modern science in the seventeenth century in England and France, with the formation of the first scientific societies. The important achievements of science studies in the 1960s and 1970s thus represent a significant step forward in the understanding of scientific activity, particularly in terms of special collaboration "forms" in science. Nevertheless, the historical coincidence at this time of continuing reduction of costs and obstacles in terms of transportation and communication, reinforced the attention given to the "collaborative" potential of scientific activities.

A second aspect of the research programme initiated in that period, and reinvigorated in the late 1980s and 1990s, is the increased attention given to the "scientific community" as an important dimension of human activity and also as a distinct subgroup within the more general social system of science and technology. These studies provided a thorough and in-depth analysis of the research community as an interesting social subsystem brought more clear evidence of the norms, values and behaviour of researchers and research groups.[14] More recent research, during the 1980s and 1990s, sought a better understanding of the complementary aspect of the "allocation of resources" within these scientific communities. This research on the "new economics of science" attempts to explain the organisation of science and the incentives underlying scientific activities, as well as the different "forms" assumed by the various actors involved in the science system, usually above the level of the

[14] In this regard the notion of a "scientific community" is extensively discussed by Hagstrom (1967) and Ziman (1984).

individual scientist.[15] Some applications of the "new economics of science" rationale on the study of research in Europe are exemplified by Garcia-Fontes and Geuna (1999) and Geuna (1999). Here again, the "research network" or the diversified "forms" of "research networks" are considered as one among several "forms" of social structures reflecting the organisation of science - and principally the collaborative nature of science.

If one discusses the "research network" concept as one form of research collaboration, the following comments are worthy of more detailed consideration. Recent research (see e.g. Katz and Martin, 1997) has highlighted the importance of reflecting on the concept of "research collaboration", the motives and incentives for the collaborative nature of science and the typology of "forms" of collaboration. Science collaboration should be looked at as a non-precise and sometimes quite fuzzy process, whose boundaries and components are not always rigorously delimited. Within and between the various levels of collaboration (individual, research group, departmental, institutional, national and international), the authors suggest a taxonomy for approximating the concept of "research collaboration" to the actual practice of science. Moreover, attention is given to the problems and limitations encountered in "measuring" forms of research collaboration using bibliometric indicators such as co-authorship (see previous section 2.1.3). These "measurement" problems are discussed later in this section.

Nonetheless, the empirical evidence corroborates the existence of collaborative structures above the level of the individual scientist, reflecting the organisation of science communities. We can find historical evidence going back to the modern science period in the seventeenth century and the formation of the "formal" institutions of science, such as the Royal Society of London in England and the Académie Royale des Sciences in France. When analysing current research and education systems all over the world, and the most typical forms of organisation, we again find the prominent "forms" of the "research group" as a social structure, agglomerating individual researchers, in a more or less tightly bound, more or less multi-disciplinary setting of research. Much more informal "invisible colleges" of collaborators, also dating back to the beginning of modern science, without any visible institutional materialisation or formal existence are also manifest throughout the history of science (see de Solla Price, 1968; Crane, 1972; and Mullins, 1972).

[15] For a survey of this literature, see David, Foray and Steinmueller (1999), and for a detailed examination of the research programme of "New Economics of Science", see Dasgupta and David, 1987 and 1994; and Diamond, 1996)

The wide diversity of processes of knowledge creation, dissemination and use led to the organisation of specific forms of scientific inquiry and the development of particular institutions for achieving scientific consensus - and more contemporarily of particular forms of "networking" of research resources. The formation, development and reconfiguration of these "network forms" of organisation within scientific communities results from this interactive and continuously dynamic process of social and epistemological negotiation among scientific actors - researchers, research groups and research institutions. These "research networks" - organisational forms constituted by research actors and their interactions - assume a wide spectrum of collaborative forms. If, on the one hand, they can be categorised in terms of their level of collaboration (from inter-individual to international collaboration), on the other hand, they can be seen as more or less formalised/institutionalised (from informal structures to completely institutionalised collaboration networks).

Research network formation, sustainability and restructuring are strongly dependent on the incentive structures for collaboration and the motives for collaborative activity. Another extensive branch of the literature has focused on explaining the factors motivating research collaboration in its diverse forms. In recognising the collaborative nature of scientific activity as well as its international character, a number of reasons have been considered as incentives for the formation and development of collaboration in science.

The growing complexity of the research problems to be tackled by science, as well as the multi-disciplinary nature of the required efforts, have been considered as key factors in explaining the agglomeration and pooling of research resources (see e.g. Solla Price, 1987).

A second factor concerns the natural evolution of the scientific enterprise, with the progressive specialisation of research efforts in more precise and detailed topics of research and sub-specialities, supporting an accumulating division of labour and with incentives for a greater degree of specialisation and more complementarity in the allocation of resources.

The increasing costs of large-scale research projects (particularly funding "big-science" research in areas such as high-energy physics, environmental protection and human-computer interaction), means that the financing requirements cannot be supported by a single institution, or even by a small set of research institutions (on the topic of Little science, Big science see de Solla Price, 1986).

Another reason for more collaboration in science is the relatively recent - mainly between the 1980s and 1990s - budgetary constraints on the total public funding of academic research and, particularly, fundamental research efforts, that sets some limits to this kind of research activity. Moreover, the opportunities brought by more accessible, faster and more reliable transportation and communication infrastructures that allow the cost effective mobility of researchers and other research resources (e.g. scientific instrumentation or scientific information).

Science policy initiatives that endorse collaboration efforts within and between sectors of research (public, private and other sectors) and fuel the innovation process in order to exploit the scientific potential for economic growth, long-term employment conditions and quality of life, constitute another incentive for collaboration (specifically with regard to science policy, see later discussions in this section).

Even if not exhaustive, the above list includes some of the arguments commonly used to explain the increase in collaborative activities. Unlike the attention paid to the advantages and incentive factors involved in collaboration, much less attention has been given to the costs and additional problems resulting from collaborative research (a welcome exception being Katz and Martin, 1997). The additional costs entailed in the mobility of researchers, the overheads caused by inter-sectoral, international and inter-disciplinary research activities, and the increased complexity in the management of different research groups with different research "styles", are some of the factors that need to be taken into account when planning research in a collaborative environment.

Some of these factors supporting the organisation of scientific activity in a collaborative form, and including the "networking" of research resources (human, material, financial and informational), are also the focus of science policy action. A case in point involves science and technology policy strategies for the formation and development of "research networks", "centres of excellence", "virtual centres of excellence" and other research "networking" forms.

In addition to the factors summarised above, and particularly those concerned with strengthening science-technology linkages to the benefit of innovation systems and society in general, some other elements are considered to justify particular initiatives of science policy in this regard. They include the following. First, there is the need to avoid duplication of research efforts at a national, regional or international level, and achieve associated efficiency gains with regard to the allocation of resources for research and development. This is generally considered a form of minimising the effects of fragmentation and balkanisation of research, particularly at the international level. Secondly, there is a need for more efficient integration of research strategies at a level higher than the national, and particularly at the continental level (this is certainly the case for European networking efforts). Thirdly, there can be optimal achievement of a "critical mass" of research resources (when this is not attainable in isolation), if a set of excellent research groups/infrastructures pool their resources. Finally, it is commonly assumed a benchmarking goal of identifying "excellence" in research for individual topics of research, and the associated inter-linking of such "centres of excellence" in order to attain a competitive position in relation to other world areas (e.g. the European Research Area, competing against Japan or the US).

These are some of the arguments for supporting the formation of "research networks", as institutionalised forms of collaboration within well-defined topics of research (naturally considered to be relevant for the regional innovation systems under analysis). However, the above does not tell us what the outcome of the new networking processes is likely to be and thus we need to introduce some "standards" into the measurement of performance and the comparative assessment of different types of collaborative activity. We briefly outline some of the associated literature in the following paragraphs.

We could recognise two distinct but complementary aspects of this problem: on the one hand, the measurement of research performance (or productivity) at different levels; on the other hand, the measurement of "collaboration".

Traditionally, and since the pioneering work of Derek de Solla Price in the 1960s, bibliometric indicators have been used as a proxy for research productivity at level of the individual researcher (individual scientist). In fact, the use of publication counts (number of refereed publications in scientific journals), impact (number of citations to a given researcher's publications) and inclusion in the list of the most cited papers in a delimited topic of research, are usually taken as approximate indicators of the scientific standing or reputation of an individual scientist. The combination of these bibliometric assessments with a "peer review" process conducted by "experts" in the field constitutes, even nowadays, an important component of evaluation procedures. Nevertheless, the same "measurement" criteria are potentially not valid for forms of research collaboration above the level of the individual - such as whole research networks. However, it is not unusual to take the aggregate productivity indicators of researchers belonging to the same "research group" to assess the overall performance of that group. Some tentative exercises, along the same line of argument, have tried to assess "institutional productivity". However, at the national or international levels, such approaches seem to be fairly unsatisfactory. More recently (see e.g Van Raan, 2000), some other procedures have been suggested for this task. These methods are based on a combination of bibliometric techniques and science-mapping network analysis.

The measurement of collaboration structures, as opposed to the performance of those collaboration structures, also suffers from some natural obstacles. The problem of the dynamic nature of these "collaboration structures", as they change over time, is tackled in a later section (see section 4.2.4), but it represents a fundamental problem when conducting non-longitudinal analyses (particularly cross-country, single-period analyses).

The use of bibliometric indicators (such as co-authorship) with the purpose of measuring collaboration has been shown to be only an approximate and partial indicator (Katz and Martin, 1997) - even at the level of groups of researchers co-authoring publications and this partiality is particularly reinforced when moving from inter-individual collaboration to levels of institutional, national and international collaboration. Nevertheless, the use of co-authorship to identify the best-connected authors and the most active and interacting researchers in specific fields has been extensive (Persson and Beckmann, 1995; Melin and Persson, 1996; Newman, 2001). The literature on the analysis of patents and linkages from those patents to research publications is extensive (*e.g.* Jaffe *et al.*, 1993, Patel and Pavitt, 1994; Cohen *et al.*, 2002; Gittelman and Kogut, 2003). However, more recently new and more advanced methodologies have been tried in order to infer "collaboration for innovation" or "networks of inventors" based upon cross-linkages of patent and publication data (see e.g. Balconi *et al.*, 2002).

More recently, other indicators of research activity (such as participation in collaborative research projects) have been used. These investigations cover analyses of research and technological development projects, as well as fundamental and applied research projects (Removille and Clarysse, 1999; Luukkonen and Hälikkä, 2000; Whidhalm *et al.*, 2001). This line of research has usually been directed toward analysis of the impact of research policy (particularly European research policy), especially at the institutional, and the national and international levels, but less so at the level of the research group.

The above literature review highlights some of the fundamental problems related to the measurement of "collaboration" at various aggregate levels. These problems are compounded when it is recognised that this is but one of the dimensions of collaborative research (namely the level of collaboration). The measurement of informal collaboration structures, i.e. the "forms" of collaboration that are not visible or institutionalised, is almost impractical, other than by surveying the research community directly.

A combination of several methods, such as bibliometric analysis, network analysis of research activity and surveys, is likely to provide better results when identifying structures of scientific collaboration, but not necessarily when assessing the research productivity of such collaboration "forms" (for mapping networks of excellence, see e.g. European Commission, 2001).

The previous sections of the literature review highlight the wide spectrum of "forms" involved in the collaborative and communicative nature of science activities. These patterns of communication (along a spectrum of formal and informal processes) and structures of collaboration (from the individual researcher to large collaborative and institutionalised efforts) are likely to change in electronic environments. However, the effective direction of this transformation is not obvious, nor is the nature of the structural change, and the extent to which new structures are likely to emerge and or replace traditional ones.

This transformation process naturally depends on the inherent dynamics of research communication and collaboration but also on the characteristics of the electronic environments available for scientific communication and collaboration. These electronic networks, and computational and communication resources are widely analysed under the common theme of "digital infrastructures for science".

There is an effective science policy interest in the establishment of these digital infrastructures for science. This research area involves several research topics. Section 2.1.5 reviews some of the literature on the establishment of electronic infrastructures for science, high-performance large-scale networking, infrastructures and services, exchange of data and large-scale database resources, access to scientific instrumentation, digital libraries and electronic publishing in science, and collaboratories. GRID systems have become representative of a new class of distributed computing that provides information and services to very large and distributed groups, as well as access to resources and distributed processing power, just like in an electric GRID (recall definition in footnote on page 3) .

The focus of this current investigation is on changing patterns and structures of communication among researchers in the transformation to electronic environments, or in other words, on the extent to which electronic environments are supporting research collaboration activities. Research on this is more extensively reviewed in section 2.1.6.

2.1.5 Electronic Information Infrastructures for Science

Investigations of the potential impact of information and computational infrastructures on the process of scientific research, collaboration and communication, span a wide variety of disciplines, as well as appearing in the literature from several different periods with an early beginning in the 1940s, extensive research in the 1960s - 1970s, and more recent research in the 1990s. We briefly review some of this literature.

Studies on the importance of computing platforms and electronic networks for the practice of research and dissemination of its results go back to the 1940s. In an article published in 1945 (*Atlantic Monthly - As we may think*), Vannevar Bush, the then director of the U.S. Office of Scientific Research and Development, provided a visionary account of the potential impact of these technologies on the way researchers gather, store, find and retrieve information. In 1965, J.C.R. Licklider published a book, entitled *Libraries of the Future*, about the numerous ways digital computing could transform libraries and information exchange (both above references are cited in *Digital Libraries*, William A. Arms, 2000, as significant early contributions to this research area).

During the 1960s there was a significant programme of research on scientific communication, initiated in 1961 by the *American Psychology Association* (APA), which involved an extensive number of studies on the effective potential of new information systems for improving the formal system of scientific publishing. This programme of research was entitled APA's Project on Scientific Information Exchange in Psychology and was jointly directed by William D. Garvey and Belver C. Griffith. These studies included a wide range of analyses: the study of the sociological nature and dimensions of the process of scientific communication and science research (Menzel, 1968); the analysis of empirical publication data for assessing collaboration structures in "invisible colleges" (de Solla Price and Beaver, 1968); the analysis of social innovations in scientific communication, from which resulted the traditional "Garvey and Griffith model of scientific communication" (Garvey and Griffith, 1968); to more technical analyses of the way electronic information services might affect scientific journals (Swanson, 1968).

The development of the Internet in the 1960s and the setting up of its first effective network in 1969, with the establishment of the DARPA project electronically connecting four research centres in the U.S. (University of California Los Angeles, Stanford Research Institute, University of California Santa Barbara and University of Utah), began a new era in both dimensions: the technological development of networking services and electronic environments and the trend to electronic scientific collaboration and communication (Leiner *et al.*, 2002). Since the standardisation of Vinton Cerf's Transmission Control Protocol/Internet Protocol (TCP/IP) protocol in 1973,[16] the progressive technological development of computing performance and capabilities as well as the development and explosive growth of network technology, have allowed a tremendous increase in connectivity of research organisations to the Network. In 1986, the decision of the American National Science Foundation (NSF) to open its backbone infrastructure to national and then international connectivity, allowed the progressive extension of the Internet worldwide, as well as its transformation from the initial restricted "Research Collaboration" setting to the public electronic network we use today (Leiner *et al.,* 2002). A third factor is relevant to the current wide availability and use of electronic networks: the continuing improvement in the computing graphical interface and the standardisation of Internet technology allowing the simple but efficient inter-linking of resources and information worldwide - this shift is usually dated to 1992-1993 with the establishment of the HTML standard and the beginning of the World Wide Web or "the Web".

These two dimensions, the wide availability of electronic networks and Internet services, and the public nature of the information disseminated in those networks, are probably two of the most important causal explanations of the recent upsurge of interest in studies on electronic scientific communication within academic communities. If, on the one hand, the wide availability of these electronic networks and associated services is a contributing factor to their more intense use within academic communities, on the other hand, the public nature of these electronic networks might impede or even block the special needs of scientific communication and collaboration, due to lack of security, lack of robustness in communication and degradation of service. Some of these issues are related to the specific requirements for very reliable and high-performance electronic networks; those needs include electronic systems for the "validation" of knowledge content and authenticity, and

[16] As an interesting coincidence, the first public disclosure of the TCP/IP protocol - the standard protocol supporting Internet services – occurred at a public conference held, in 1973, at the University of Sussex – the place where these lines are being written.

legal issues related to property rights, and quality of service comparable with the traditional scientific collaboration systems.

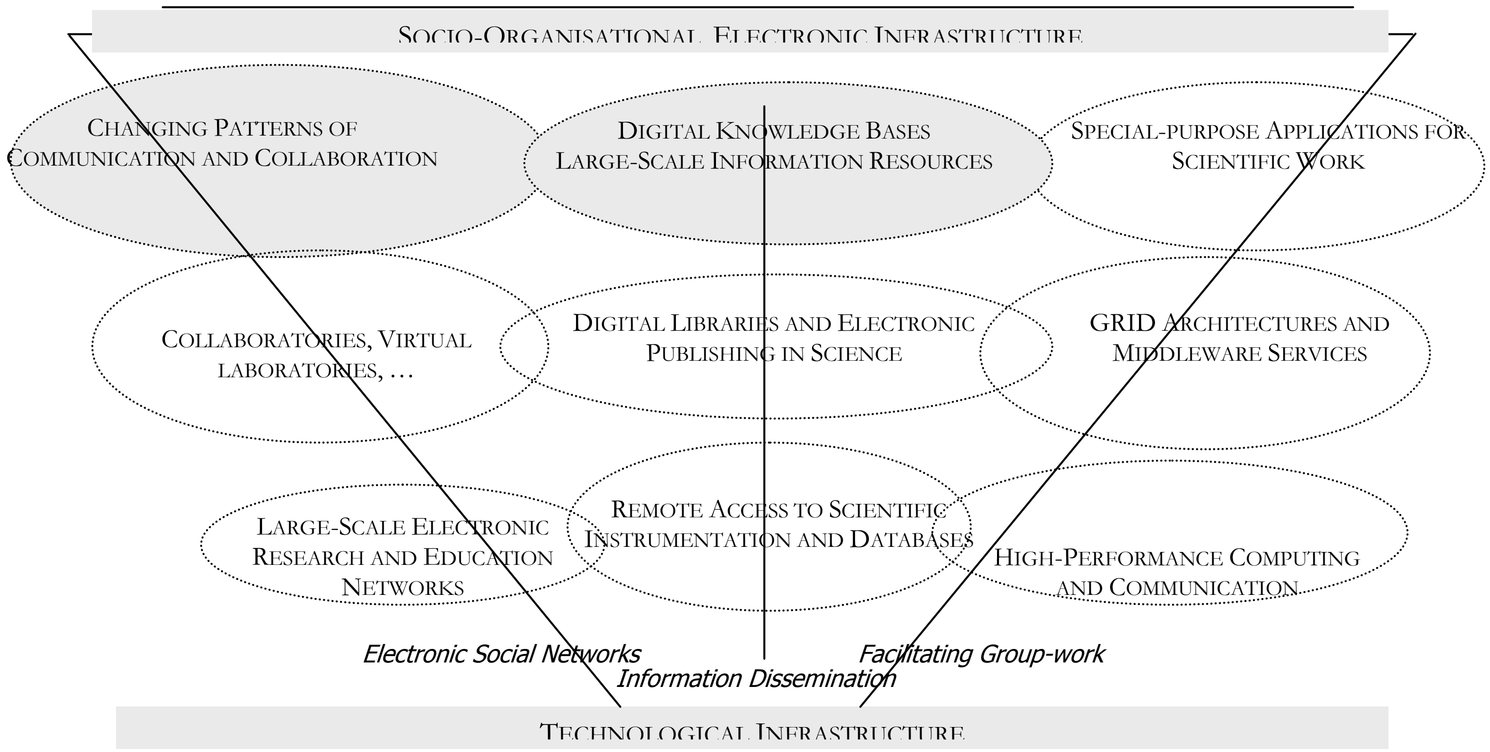
Above, we have historically situated the theme of the impact of ICTs and electronic networks on science. The following discussion in this section is focused on the detailed analysis of the various research topics under the common label of "digital infrastructures for science".

A significant problem area in science policy investigation has been the changing nature of the science system under the effect of advanced ICTs. An integrated perspective would consider the following topics: the issues related to the establishment of digital infrastructures particularly effective for scientific activity - whether physical networks or scientific applications; the problems related to access to scientific information; the analysis of changing patterns of communication among scientists; the technological questions related to the availability of remote scientific instrumentation; the revolutionary area of electronic publishing and digital libraries; and the impact of ICTs on education and training, to name but a few of the most important research problems (see Aubert *et al.*, 1999; OECD, 1998; and OECD, 1999).

The literature review is discussed within a conceptual framework, distinguishing two complementary dimensions: the *technological nature* of these digital infrastructures and the *application nature* of these digital infrastructures. The technological nature of these digital infrastructures is examined from two different aspects: either in terms of being more oriented towards analysis of technological systems or being more oriented towards socio-organisational systems. The *application nature* of these digital infrastructures for science is analysed in terms of being more oriented towards the facilitation of group work, the support for a division of labour and collaborative activity - electronic social networks - or the distribution of information and other electronic resources - information dissemination.

All of these dimensions are important research topics *per se*. Significant research has been directed towards some of them. Figure 2.1 is a conceptual map of the various topics of research under the general theme of "Digital Infrastructures for Science".

Figure 2-1 - Conceptual map of topics of research in“Digital Infrastructures for Science”



Science policy initiatives have been focused essentially on the more technologically oriented aspects of the digital infrastructure, as well as on the creation and development of technologies facilitating group-work and allowing the distribution of electronic information and electronic resources. In coming years the electronic applications supporting socio-organisational systems are likely to gain increased significance. We start the analysis by discussing the literature on the more technologically oriented infrastructures and in the final part of this section and the following section, move to a detailed analysis of the more socio-organisational applications of these electronic systems (particularly electronic networks supporting socio-organisational systems, thus forming electronic social networks).

Studies of the influence of electronic technologies on scientific communication have mainly concentrated on the provision of electronic physical infrastructures (Axmann and Axmann, 1999), the use of e-mail in communication among scientists (e.g. Walsh and Bayma, 1996; King and Mckim, 1997), collaboratories and "extended research groups" (e.g. Finholt and Olson, 1997; Carley, 1996), issues of electronic publishing and digital libraries (Arms, 2000; Roosendaal, 1999; Lawrence *et al.*, 1999; Ginsparg, 1994, 1997), and more recently computer-supported social networks in on-line environments (Wellman, 1996; Garton *et al.*, 1999; Koku *et al.*, 2000).

Science policy action has been recognised to play an active role in the transformation to electronic scientific infrastructures. Attempts to maximise the benefits of advanced ICTs for scientific research and collaboration have come to the fore in recent science policy initiatives. For example, in its last Comprehensive Spending Review, the UK government announced a financial commitment of £125 million for new e-science projects. The 2001 White Paper on Science and Innovation signalled the government's intention to support new digital infrastructures in science: there is 'an ever-increasing need for exchange of complex data…[t]o ensure that the UK remains at the leading edge of network provision, we will improve the academic high-speed network' (Byers, 2001: 14). The UK government has also linked developments in this new digital infrastructure to further economic opportunities: '[t]his new generation of e-science infrastructure will underpin the next generation of e-business technology world-wide later this decade (Byers, 2001: 21). Commentators on science have also pointed to the role of new digital infrastructures in reshaping communication. Gibbons *et al.* (1994) suggest that the use of new ICTs is helping to restructure the nature of knowledge production. David and Foray (1995) argue that the new electronic exchanges have the potential to alter the traditional norms of scientific behaviour.

Research on electronic networking infrastructures and high-performance computing and communication spans a wide diversity of topics. In a recent report (PITAC, 1999), the US President's Information Technology Advisory Committee made important recommendations with regard to a more scalable information infrastructure in order to cope with a larger number of users, users demanding high reliability and short periods of loss of service and even mobile users requiring rapid reconfiguration of networks. They also stressed the fact that in order to support high-end computing in conducting scientific research there is a need for extremely fast computing systems, with both rapid calculation and rapid data movement. In the US, the Next Generation Internet (NGI) Initiative (which began in October 1997) constitutes a large-scale testbed for advanced networking infrastructures (NCOITRD, 2001). The NGI initiative is a multi-agency research and development programme that is developing both advanced networking technologies and revolutionary applications that require advanced networking.

The provision of advanced physical electronic networks allowing the connection of local/regional area networks to very high-speed backbones of the (inter) national research networks (Axmann, 1999) is underlined as an important science policy initiative with regard to European electronic networking. In the same vein, a joint statement between the Academia Europaea and the European Science Foundation reinforced the need for high-bandwidth computer-based networking in Europe (ESF, 2000). Research has also been directed to detailing networking research needs (IWG, 2001). Among these, there is substantial research directed towards building adaptive, dynamic and smart networking infrastructures, the measurement, modelling, simulation and scalability of networks, issues of trust, security, privacy and reliability, networking applications and networking middleware (in the same direction of GRID architectures). Another area of investigation has been the enhancement of the public information infrastructure for science, leveraging open and public knowledge and information resources (NIST, 2001). Important topics in this regard are the non-commercial infrastructure, which is in place and evolving, the user orientation approach, which implies less focus on digital collections and is more about providing services to a community of users, and the opportunities for leverage of resources through an open infrastructure.

In analyses of the future of information infrastructures for the physical sciences (DOE, 2000), a number of prerequisites have been identified as essential for realising the potential of ICTs for science, among which is the networking information infrastructure that has to be more than a storage and retrieval capability. Rather, in the long term, it has to support fundamentally new ways of doing science (DOE, 2000: 1-3).

The more integrated use of some of the above technologies and other electronic environments has recently received significant attention. The specific needs of research communities in terms of special dedicated computational resources, as well as restricted electronic networks, have been subject to intensive science policy discussion and effective action. The most visible of these science policy efforts have been concentrated on the implementation of "National Research and Education Networks", the conceptualisation and implementation of "Collaboratories" and the technical implementation of "GRID systems" as distributed systems providing resources just like in a traditional electric power GRID. These technological infrastructures for scientific activity and research collaboration are briefly summarised.

Some of the most focused efforts in the development of specific electronic infrastructures for science have taken the form of National Research Networks and the international interconnectivity of those national research and education communities over transnational electronic backbones (Sabatino and Arce, 2000). Worldwide these infrastructures are usually high-bandwidth electronic networks, providing more reliability and quality of service, allowing access to specific scientific resources (particularly to electronic networks, such as the Internet, but also to specific documentation resources and technical support), and connecting university-wide area networks and other education facilities (UCAID, 2002; DANTE, 2002). On top of these electronic infrastructures, access to remote scientific facilities, and other more application oriented services, such as "collaboratories" and Grid systems are likely to emerge.

Research on "Collaboratories" dates back to a White Paper exploring the concept of a national collaboratory by Dr Wulf, while he was assistant director of the National Science Foundation's Directorate for Computer and Information Science and Engineering (Wulf, 1989). He coined the term "collaboratory" by joining the words "collaboration" and "laboratory". The concept is defined there as a center without walls, in which the nation's researchers can perform their research without regard to geographical location - interacting with colleagues, accessing instrumentation, sharing data and computational resources, and accessing information in digital libraries. In the future, science and education are likely to function in a highly distributed and fully connected environment, where each resource or agent can interact with every other. This vision of Collaboratories has underpinned various initiatives for the development of collaborative technology. In the US, programmes such as DARPA's Intelligent Collaboration and Visualization program and the Department of Energy's DOE 2000 initiative for national collaboratories, constitute very good examples. The concept of a "Collaboratory" is extensively discussed in Finholt and Olson (1997) and Olson (1999 and 2001). In CNC (1993) a detailed examination of collaboratory research is provided with extensive discussion of already functioning collaboratories for Oceanography, Space Physics and Molecular Biology. A survey of some other functioning collaboratories is provided in Interactions (1998). [17].

GRID technology (see Smarr, 1999, for a general overview of GRID systems) has been seen as the technological successor of the world wide web. As the WWW has provided easy access to information, GRID technology is likely to provide similar easy access to high-performance computing power, data processing and sharing of resources. The "Grid" is a set of pervasive technologies that allow the sharing of knowledge resources in large-scale electronic networks with excellent Quality of Service. National "e-Science" programmes (see e.g. the UK e-Science initiative in Atkinson *et al.*, 2001) are currently endorsing the adoption of GRID technology and its applications in several scientific communities. In the US the standardisation of GRID technology has also been a strategic policy objective. At international level, a standardisation process similar to the highly successful Request for Comments (RFC) system of Internet technology, has been recently set up - see the Global Grid Forum Documents and Recommendations (GGFDR, Catlett, 2001).

Another line of investigation has focused on access to and wide dissemination of digital libraries and reliable electronic transfer of scientific data.[18] Digital Libraries are usually defined as managed collections of information widely disseminated over large-scale electronic networks. In PITAC (2001) a vision of digital libraries as universally accessible collections of human knowledge is put forward along with a detailed discussion of the major challenges posed to the future of digital libraries. All citizens anywhere anytime can

[17] The *Space Physics and Aeronomy Research Collaboratory* (available at http://www.si.umich.edu/sparc), the *Great Lakes Center for Aids Research* (available at http://www.greatlakescfar.org/), the *NSF-funded Medical Collaboratory* (available at http://www.si.umich.edu/medcollab/), or the *Collaboratory for Research on Electronic Work* (available at http://crew.umich.edu/).

[18] See e.g. NRC (1997), for general overview of international open exchange of scientific data and Arms (2001) for a general overview of Digital Libraries research.

use any Internet-connected digital device to search all of human knowledge. It is stressed in this report that new libraries offer digital versions of traditional libraries, museums, and archive holdings, including text, documents, video, sound and images, but they also provide powerful new technological capabilities, enabling users to refine their enquiries, analyse the results or change the form of information to interact with it.

The US has been leading advanced research on digital libraries. Phase I of the Digital Library Initiative funded 6 projects of about $6 million per year, from 1994 through to 1998. Six university-led consortia[19] conducted R&D in applying advanced computing and networking capabilities to make large distributed electronic collections accessible, interoperable and usable (PITAC, 2001). Phase II of the Digital library Initiative, started in 1998 and is funded at the level of about $11 million per year, having supported more than 37 projects up to 2002. This second stage of the Digital Library Initiative aims to extend research and testbed activities in promising digital libraries areas, to accelerate the development, management and accessibility of digital content and collections, and to encourage the study of interactions between humans and digital libraries in various social and organisational contexts (DLI, 1998).

A survey of digital library projects was provided in a Special Issue on Digital Libraries (IEEE Computer Magazine, 1996). Research issues in relation to digital libraries include: acquiring and storing information, finding and filtering information, securing information and auditing access, universal access to large collections of information, cost management and financial instruments and socio-economic impact (Adam, 1996). Topics for investigation span more general research on the concept, the underlying assumptions and characteristics of digital library systems (Fox *et al.*, 1995; Levy and Marshal, 1995; Wiederhold, 1995), digital libraries usability, user interfaces and user studies (Sumner and Dawe, 2001; Börner *et al.*, 2002; Bishop, 1999) and technical issues related to document preservation and knowledge dis-aggregation, the interoperability of systems, and collaboration, multimodality and plurality (Bishop, 1998; Hedman, 1999; Blanford *et al.*, 2001). Other more specific lines of research have focused on world-wide access to digital libraries and resolving the digital divide (PITAC, 2000); and the specific application of

[19] The six Digital Library Phase I projects were: University of California at Berkeley (Environmental planning and Geographical Information Systems); University of California at Santa Barbara (The Alexandria Project: Spatially-referenced Map Information); Carnegie Mellon University (Informedia Digital Video Library); University of Illinois at Urbana-Champaign (federating repositories of Scientific Literature); University of Michigan (Intelligent agents for information location); Stanford University (Interoperation Mechanisms among heterogeneous services)

digital library technology for Science, Mathematics, Engineering and Technology Education (Agogino, 1999).

An area of research closely inter-related to research on digital libraries is electronic publishing in science and the changes likely to occur in the scientific publishing system. In addition to the research reviewed above, a book edited by Eckman and Quandt (1999) provides an extensive analysis of the several dimensions involved under the issue of electronic publishing, digital libraries and the scholarly environment. These dimensions encompass matters of technology (scholarship data needs, digital image quality, content licensing), empirical studies of electronic publishing (e.g comparing electronic journals to print journals and analysing pricing-models for e-publishing systems), empirical studies of the usage of e-journals (scholarly practice of access to on-line journals, patterns of use of electronic journals, and consortium-type access versus ownership) and finally visions for the future of electronic journals. In an analysis of the future for electronic journals, Varian (1999) provides an economic analysis of the potential evolution of electronic publishing (seeking to understand the market model underlying both systems - the traditional print journal and the new electronic journal.

Much research has been conducted on systems of electronic publishing other than electronic journals (e.g e-print archives and pre-print servers). In this regard high-energy physics constitutes a pioneering example. Starting from 1991, automated archives for the electronic communication of research information have been serving tens of thousand of users world-wide (Ginsparg, 1996). Most of the functions of the traditional publishing system are now available in these electronic archives (in terms of content, functionality, methodology and appearance) with additional advantages (e.g. in terms of automation, dramatically accelerated time-scale, document structure and additional data management and presentation functionality). Roosendaal and Geurts (1999) provide a policy-oriented analysis of the developments in electronic publishing and the likely changes in the function of the several elements intervening in this system. They analyse the traditional functions of awareness, certification, registration and archiving, as well as the various actors (readers, authors, publishers and librarians). They conclude by analysing essential organisational changes in the system.

This section has summarised some of the research on digital infrastructures for science, with a particular focus on the distinction between how these technologies are analysed essentially as technogical systems, on the one hand, or as socio-organisational electronic infrastructures effectively supporting the work and collaboration of social networks, on the other. The extent to which these technologies are regarded as mere technological systems might damage the potential benefits for science. In fact, the endogenous characteristics of science in terms of collaboration and community organisation (as identified in previous sections) require a more sociological and transactionist approach to the adoption of these electronic systems, in order that they support the traditional structures of collaboration as well as helping to create more effective and innovative forms of electronic research organisation. A second and complementary dimension framing the previous discussions is the detailed examination of the *application nature* of these technologies for science purposes. If the impact of these electronic networks is restricted to the replication of traditional forms of work organisation or group-work, then the overall effect on the organisation of the science system might be minimised (this might result from particular science policy inititatives widely implemented in this direction). On the other hand, if more extensive policies and initiatives, as well as the natural evolution of the adoption of these technologies by scientific communities, is oriented towards the public and wide dissemination and distribution of electronic resources (data, information and knowledge), as well as the implementation of infrastructures facilitating electronic collaboration, and electronic social networking, then the overall benefits for science might be heightened.

This research particularly focuses on the question of changing patterns of scientific communication with the availability of electronic infrastructures, and particularly how scientists and practitioners *use* Internet technologies for scientific work, communication and collaboration. However, this question is closely linked to studies of the social organisation of science, studies of research collaboration and particular forms of collaboration such as research networks. This social context of science strongly influences the degree to which technological opportunities may facilitate the communication process in science.

In recent years, renewed interest in electronic networks has led to theoretical and empirical investigations of electronic scientific communication and collaboration. This research has been mainly directed towards a better understanding of the science policy implications, and a broader conceptualisation of the scientific communication process, as well as the sociological analysis of electronic scholarly communities. This literature is briefly reviewed in the next section.

2.1.6 Electronic Scientific Communication and Changing Patterns of Collaboration and Communication

As the contribution of this investigation is particularly focused on achieving a better understanding of the electronic networking of social networks in science, this section analyses in more detail studies of e-science with a specialised focus on the interaction between the technological characteristics of electronic networks and the socio-organisational nature of scientific communities. Previous sections analysed in detail features of the communication process in science (sections 2.1.2 and 2.1.3), as well as features of collaboration and forms of organisation of scientific communities (section 2.1.4). The previous section integrated the analyses of digital infrastructures for science in terms of the technological nature of these technologies (and the extent to which they support socio-organisational systems) and the characteristic of the scientific applications being implemented (in terms of facilitating group-work, allowing the disseminaton of information and supporting social networks of electronic collaboration). In this section we come closer to this latter topic, by approaching electronic networks essentially as social networks and analysing the use of electronic methods for transactional purposes.

Research work on the interaction of computer networks with scientific workgroup organisation further testifies to the systemic relationship between the use of ICTs, and the social structure of the scientific community. In fact, as long as computer networks are used to link scientists or groups of scientists they become more like "social networks" than technological networks (Wellman, 1996). If these electronic networks of scientific interaction are to be considered as social networks, then we may go a step further than the traditional investigations along the lines of Computer Supported Cooperative Work (CSCW), and Computer Mediated Communication (CMC) (for a survey see e.g Koku, 2000). Both these classes of investigations attribute to the technological environment a decisive role in the interaction with work processes and information exchange. On the other hand, if the intertwining of computer networks and, electronic environments more generally, and scientific social networks is taken to a higher level of interaction, then we may take advantage of a whole set of techniques and methodologies for analysing these systems of electronic social interaction.

These works suggest that many of the features of science are subject to the influence of new technology and suggest that the impact of the new technologies often depends on the development of a new set of social institutions that re-order social relations to take account of the character of the technologies.

In a line of investigation that follows alternative methods to this investigation for conducting empirical analyses of e-science some empirical analyses (supported by user surveys of scientists from different disciplines - mathematics, physics, astronomy, chemistry, experimental biology) have shed light on the transformation of scientific communication as we move into digital infrastructures (see Crawford *et al.*, 1996; Garton *et al.*, 1999; King and McKlim, 1997; Walsh and Bayma, 1996). Some seeming consensus emerges from these investigations. First, there is need for a change in the conceptual representation of the traditional Garvey/Griffith model of scientific communication - which described the communication process in science in too linear a way, from the scientist's initial conception of an idea and its preliminary reports through the presentation of preliminary findings in conferences, until final publication in journal articles and its acceptance as reliable and stable scientific knowledge by the scientific community at large. Secondly, the intensity and pattern of use of ICT for scientific communication varies significantly across scientific disciplines and even among sub-disciplines. Some mature results corroborate the hypothesis that more geographically dispersed and interdependent communities, and research relying heavily on expensive and large experimental research settings (such as experimental physics), benefit more from electronic communication. The more application-oriented fields (such as chemistry and experimental biology) with frequent commercial interests, are less amenable to the use of ICT for informal scientific communication, depending much more on formal publication or patenting activities. Thirdly, the inter-disciplinary nature of the research field affects the collaboration patterns of the research community and, as such, the overall communication process. In this regard, several other empirical studies have analysed the impact of information technologies on specific and interdisciplinary fields of research. Such is the case of materials science (Schwartz *et al.*, 1996), engineering (Donohue, 1996), and earth sciences (Heming, 1996).

Some other sudies develop methods and a conceptual orientation similar to those followed in this investigation. In these more sociologically embedded endeavours, research on the use of computer networks as a basis for supporting social networks of researchers (Wellman, 1996) is significantly different from studies of CSCW and CMC. In the latter research areas, it is the design of the technological environment that shapes the work process and content - work occurs online. By contrast, in scientific networks, work processes are primarily conducted offline and the references are to common knowledge of scientific ideas, findings, or equipment. While it is possible that future developments of *collaboratories* and other means of direct interaction will lead to a convergence of these fields, at present the principal focus of research is on the structure of these scientific community interactions. In this respect, more generic methods for studying the formation of social networks in online environments are relevant (Garton *et al.*, 1999).

Some other studies that reinforce the contribution of this investigation include a final branch of the literature that has focused explicitly on the study of the Internet as a large-scale electronic network and an environment for community formation and information dissemination.

Originating from a research project about four decades ago in the 1960s, the Internet has become a huge repository of information and a pervasive electronic environment for communication. The exponential growth of the Internet (its actual size is impossible to determine, but it is estimated to carry more than two billion documents and to be growing at a rate of about 10,000 documents every day - see ShowDown Search, 2002) has led it to become the most complex information system ever constructed. Moreover, the decentralised architecture of the system, wherein millions of distinct individuals and organisations publish and change autonomously new information in a completely uncontrolled manner, has resulted in its becoming a highly disintegrated distributed system. Problems of slow access to information given the complexity of the system, and "information overload" given its distributed architecture, are commonly referred to as blocking factors curtailing its efficient and effective use for professional research work.

Nevertheless, recent research has highlighted the self-organising nature of the Internet (Kleinberg and Lawrence, 2001; Flake *et al.*, 2000; Albert *et al.*, 1999) and the strong regularities entailed by its structure and growth behaviour. If one considers the Internet as a sparse, highly disconnected network, where nodes (documents) are hyperlinked to other nodes, then we are able to analyse patterns in the structural organisation of the system. This research has been extended by empirical analysis to identify "web communities" and by investigations seeking a better understanding of the internal structure of the Web, and portions of it.

The crucial question for our current research purposes is whether the structure of the Internet, and particularly a restricted and bounded portion of this electronic network, reveals significant characteristics of the institutions involved, and of their collaboration patterns. Hence, we will focus our attention on the sociological explanation of the structure of electronic networks, in order to gain an insight useful for science policy. Moreover, as we analyse research networks, our primary interest is in scientific collaboration structures and the communication process in science.

Previous investigations in this area span several fields of research, including bibliometric methods for the analysis of scientific literature, sociometric analyses of social groups and institutions, and analytical methods for network analysis, to the more technological-oriented fields of information science - such as database mining and webmining, information storage and retrieval, and search-engine technology more generally.

As section 2.1.3 extensively reviewed bibliometric and scientometric studies, we focus here on the other kinds of literature.

Network analysis studies go back to the 1930s, with the pioneering work of Moreno in sociometric analysis (Moreno, 1932). The "network approach" looks at social and technological systems from a structural perspective. It identifies *relationships* between sets of *actors or entities* and measures the pattern of those relationships. Very robust analytical methods have been developed in order to measure the strength of the relationships as well as to determine the connectivity of the networks under investigation, or the centrality of some actors within the whole network. Most of these techniques are surveyed in Wasserman and Faust (1994).

In considering the Internet electronic network as a set of nodes (documents) linking to other nodes (documents), we can represent these systems as graphs and thus measure common properties of these graphs - e.g. size of the graph, the shortest path between any two documents in the graph, the overall density of the graph, the better connected nodes in the graph, or the most central and prestigious ones. In this way, we are exploring the link-structure or topology of the network in order to better understand its structure and organisation. Nonetheless, there is a crucial problem related to the highly disconnected nature of the Internet network, as well as its huge size. This leads to problems similar to those encountered with complex database systems and how to "discover" particular pieces of information in such environments.

The field of database management has dealt with problems of information classification and data-mining for several decades. The relational model of database construction has become relatively standardised as the most common method for organising *structured* information. The real problem comes with the management of *un-structured* information, such as *html documents* and other internet resources. In addition to the problem of managing a huge volume of information, we have to access, store, and filter completely unstructured data. This problem has led to the emergence of new concepts, methods and techniques particularly applicable to the "web" environment.

The field of *Web mining*, whose origins are quite recent (Kosala and Blockeel, 2000), is divided into three complementary areas of investigation: web content analysis, link structural analysis and user analysis. *Web-content analysis* is focused on the content of documents distributed in these electronic networks. Techniques and algorithms have been developed for information filtering and classification systems - particularly automatic classification systems - for categorising information in these highly distributed networks. *Link-structure analysis* is concerned with the topology of these electronic networks - with how documents are related to each other and how the whole network is structured and organised. In this area, very robust algorithms have been developed in order to explore the topology of portions of the web. Based on local linkages of subsets of the Internet, and particularly combining this link structure with topical search methods, research has indicated that one can identify highly connected components of the larger graph. *User analysis* focuses on discovering patterns of behaviour in the use of electronic resources and particularly the web. Automatic techniques have been developed in order to classify those behaviour patterns into categories that are manageable to allow the predictability of behaviour in the use of Internet resources, e.g. how many consecutive clicks a user is expected to make in order to navigate in a certain category of web sites or, on average, how long each user keeps navigating a collection of related web sites.

In this investigation we will be using techniques of link-structure analysis for determining the electronic connectivity of research institutions and the centrality of some institutions in the overall network, as well as devising new models for discovering knowledge resources on these sparse and disconnected networks. This is closely linked with search-engine technology applications as well as recent studies exploring the topology of the Internet to identify web communities. We will examine these two topics before moving on to a more sociological exploration of this problem.

Search-engine technology is mainly concerned with providing a solution to two related functionalities: *precision/relevance* and *recall*. The results of a search on the Internet should be as precise as possible in order to be effective. From the huge, and perhaps impossible to determine, collection of information available, it should provide in a few seconds more precise results on the topic being searched. This goal is sometimes at odds with the number of items of information that it can give as an output to the user (the recall capacity of the search engine). We would expect that the *more* results returned, the better. In other words, search engine technology should simultaneously provide more items and very precise items of information as an outcome of any search.

Considering the huge size of the web, and its dynamic nature, this is a very complex technological undertaking. Most search engines index documents in a database recovered from systematic crawls of the web. The crawls explore the link structure and hypermedia nature of the Internet in order to jump from one document to related documents. Complementary content analysis provides a classification of the retrieved documents. How ever refined search-engine technology has come to be, estimates suggest that 12 search-engines taken together index less than 50% of the whole internet system (Lawrence and Giles, 1999). As noted by Aquino and Mitchell (2001), search-engine technology has used three different types of algorithms to build these indexed databases: the Naïve Bayes model, which focuses on topic-word frequencies; "maximum-entropy" algorithms, which focus on word combinations and how frequently they are associated; and, perhaps the most promising approach, the "co-training" model, which studies the information on a web page, as well as the linked pages, building an association of correlations.

In fact, these strategies of combining the content of Internet resources with the link structure of those resources are better suited for "entity extraction", or the ability to build databases from collections of specific entities. This brings us closer to the notion of "web communities" and the self-organisation of the Internet. Considering a web community as a set of web pages that link (in either direction) to more web pages in the community than to pages outside of the community (Flake *et al.*, 2000), it is important to be able to identify such subsets of the large electronic network. Several studies have empirically identified such communities based on a combination of several algorithms. One of the most common ones is the HITS algorithm (Kleinberg, 1997) which explores the link structure of the Internet, starting from a set of seed URLs and determining the *Hubs* and *Authorities* resources in Internet web space. *Hubs* are Internet resources that link to many authoritative pages in the topic, while *Authorities* are Internet resources that are linked by many Hubs. There is a self-recurring mechanism in the identification of Hubs and Authorities. This problem has been overcome by refinements to the initial algorithm. The method is based on partitioning of the initial graph into well defined components. The determination of Authoritative pages on a particular topic is quite useful when ranking results of search queries. The PageRank algorithm, implemented in the search-engine Google, uses a very refined version of the initial HITS algorithm (Brin and Page, 1998; Huang, 2000). This is important to note here as we are going to use Google technology for determining our collection of *Related pages* to our initial research network (see methodology chapter, and particularly section 3.5.2).

From a technical point of view, theoretical and empirical research has demonstrated the effectiveness of using the link structure of the Internet to identify subsets of the larger network that can be categorised as "web communities". However, analysis of the inner structure of these "web communities" is still a challenge, and probably the solution is not particularly technical in nature but more sociological. In fact, we should try to understand the organisation of these electronic communities by comparison with other structural characteristics, such as the similarity of the information under exchange within these communities, or the collaboration structure of the institutions participating in these communities. This brings us closer to the specific topic of this investigation.

Sociological explanations of on-line virtual communities have focused particularly on user studies of the behaviour of those communities when participating in those electronic environments (Wellman, 1996; Garton *et al.*, 1999; Koku *et al.*, 2000). Much less research has been done on the automatic examination of electronic collaboration structures, as given by Internet resources and archives, at a level higher than the individual, and not focused on particular events (such as electronic conferences, or newsgroups participation), but unrestricted from the point of view of having a particular causal factor for their occurrence.

Under the theme of changing patterns of electronic research communication, some interesting research topics remain to be more thoroughly and empirically examined. Among these topics are the examination of the effective use of different Internet technologies for communication and collaboration activities, the intensity of use of different technologies for different aspects of the research process (and whether this is significantly correlated with scientific productivity and "seniority" in research), the extent to which traditional structures of research collaboration are being reproduced in electronic environments, the public dissemination or not of scientific resources, and the intensity of use of specialised infrastructures for scientific electronic collaboration, such as *collaboratories* and *Grid* architectures.

The next section details the research question and theoretical hypotheses underlying this whole investigation, which is attempting to fill the gap in knowledge surrounding some of these research topics.

2.2 Research Question and Hypotheses

2.2.1 Are Traditional Structures of Scientific Communication Being Reproduced in the Electronic Infrastructure of Science?

Following the previous review of the literature, in this section we identify the main gaps in that literature and highlight the fundamental contributions of this investigation. The discussion of the literature in the previous sections has underlined the importance of analysing the impact of Internet technologies and electronic networks (more generally ICTs) in terms of their effect on the organisation of scientific communities, the support of collaborative activities and the creation of innovative modes of communication and interaction. Previous studies have pointed to the importance of approaching these transformations in terms of analysing the mutual influence between, on the one hand, the new technological capabilities of these technological infrastructures and, on the other hand, the socio-organisational systems typical of scientific communities (such as features of the scientific communication process, both formal and informal, patterns of research collaboration, or particular forms of "networking" and interaction of research actors and research resources). Finally, a detailed analysis of previous investigations into digital infrastructures for science have emphasised implementation of technological infrastructures and analysis of computer-supported group-work, to the detriment of analysis of the more socio-organisational nature of these electronic systems. Much less attention has been given to analysis of these electronic networks as effective systems of social interaction and infrastructures supporting social networks of collaboration and interchange of information. This justifies a closer analysis of this particular dimension of e-science - electronic social networks - as the main focus for theoretical and empirical investigation.

Within this perspective, three main gaps can be detected in the literature, and constitute the particular focus of this current investigation. First, there is an insufficient understanding of the patterns of use of Internet technologies for research work, collaboration and communication, under the more general framework of seeing electronic networks as social networks. Secondly, there is insufficient theoretical and empirical evidence of patterns of electronic interactions and how similar these electronic patterns are to non-electronic forms of research communication and collaboration. Thirdly, there is a lack of detailed analysis of the structure of these electronic networks (for example, in terms of connectivity of research resources, dissemination of information, and centrality of certain institutions).

In the analysis of a conceptual framework pursued in the previous section (section 2.1.5) concerning the implementation of digital infrastructures for science (facilitation of group-work, distribution of information and electronic support of social networks of collaboration) we identified the need for more detailed examination of the three research issues to be empirically examined in this thesis.

Research questions

It is expected that the investigation will provide empirical evidence on patterns of communication among scientists through the use of advanced information technologies and electronic networks. The new evidence will provide the opportunity to examine how and to what extent new ICTs and particularly Internet technologies can help to reconstitute traditional norms of scientific behaviour and communication and reproduce non-electronic structures of research collaboration in digital environments.

The fundamental research questions are thus What is the structure of electronic scientific communication? Is this structure reproducing or even changing traditional (non-electronic) patterns? If so, to what extent is this transformation occurring?

The long-term analysis of regularities in patterns of use of these technologies for purposes of research communication and collaboration allows one to characterise the "structure" of electronic scientific communication. In order to be able to give affirmative answers to these fundamental research questions, analysis of usage patterns should reveal three fundamental properties. First, regularities in the use of different technologies for different aspects of research work, scientific communication and collaboration are likely to occur, even if there is some variability in the usage patterns of information and communication technologies among individual researchers. Moreover, the multi-modality of the use of the same technology (e.g. e-mail or newsgroups) in a wide spectrum of communication (formal or informal) and collaboration (individual or collaborative) activities is likely to constitute the norm. Nevertheless, the extension of this multi-modality varies significantly across technologies. Secondly, traditional structures of non-electronic scientific communication (e.g. connectivity, specialisation, hierarchy) should be reproduced in electronic environments. These structural characteristics encompass among others typical features of the science system, such as the "Matthew effect" and "Lotka's Law", the division of labour and specialisation in the structural evolution of communication and collaboration, and diversity in patterns of connectivity among researchers and research groups/institutions; as well as the subgroup structure of collaboration, reinforced or not by science policy networking initiatives. Thirdly, research collaboration patterns of connectivity should be extended into electronic networks. This being so, research results in various forms of knowledge resources are likely to be publicly available and used within these large-scale computational and communication networks. This self-organisation of electronic networks would allow the identification of "digital knowledge bases" - not electronic spaces, which constitute a mere repository of information, but an organised and managed collection of digital resources resulting from the interaction of highly connected collaboration groups.

Sections 2.2.2 through to 2.2.4 discuss in more detail these three theoretical hypotheses, which will be empirically tested in subsequent chapters.

2.2.2 Electronic Scientific Communication is a Multidimensional Process Multimodality in the use of Technologies is the norm

Traditionally, research on the analysis of usage of ICTs for research work has focused on the significance of particular technologies for particular aspects of research work (particularly e-mail, given the relatively high intensity of its use for scientific communication compared with other technologies and services). However, this is likely to represent only a small part of the spectrum of communication and collaboration activities affected by electronic infrastructures and services. This point is even more true of the more recent availability of highly distributed electronic networks such as the Internet. The following theoretical hypothesis highlights this, by assuming a different approach to these problems.

Hypothesis One: *There is multi-modality in the use of information and communication technologies for different stages of research work, communication and collaboration.*

If different technologies are hypothesised to be used in specific ways for different aspects of research work then we need more extensive empirical evidence on the use of Internet Technology for research collaboration and communication. Technologies more suited to personal inter-change of information (e.g. E-mail), while being widely regarded as relevant for most activities involved in the research process, are likely to be relatively more important for "active" processes, such as co-authoring, research work on projects, exchange of results with close collaborators and the organisation of seminars and conferences, but less relevant in other more "passive" activities such as accessing information about funding, research work by others or on-line publications. Other technologies regarded as important for various research work tasks, but more passive from the viewpoint of individuals (e.g. the Web) are likely to be considered as basic electronic information resources but not important for personal interchanges.

Assuming hypothesis one, of *multi-modality,* is true, then an enlarged set of electronic technologies and a wider spectrum of communication and collaboration activities - from the more formal to the informal activities, and from individual to highly collaborative activities - should be systemically analysed. The confirmation of this hypothesis requires the following conditions to be met.

First, while variability among individual researchers in the use of technologies is likely to occur, regularities in usage patterns of different technologies for different stages of research work, communication and collaboration should also be identified.

Secondly, usage patterns of scientists should reveal that different technologies are used more intensively and regularly for different dimensions of the wide spectrum of communication and collaboration activities. For example, we are likely to find pre-prints more suitable for "formal" electronic communication than newsgroups; hence, newsgroups are more likely to be located in the "informal" end of the communication spectrum. On the other hand, newsgroups are likely to be more "collaborative" than pre-print servers, and as such are located further along the spectrum of collaborative activities. As a second empirical example, file transfer protocol (FTP) technology is likely to be more personal (individualistic) than "collaboratories" or other advanced technologies for electronic remote collaboration. In terms of "formality" of the communication process, the relative positioning is dependent on the nature of the application.

Thirdly, and this condition partly contradicts the second one, the extension of multi-modality varies across technologies. In other words, the same technology (e.g. e-mail) might cover a wider or narrower spectrum of formal/informal communication activities, and a wider or narrower spectrum of individual/collaborative activities.

If the empirical evidence confirms the theoretical hypothesis, then the set of information and communication technologies (particularly Internet technologies) supports a wide spectrum of research collaboration and scientific communication. This wide spectrum might be structured along three complementary dimensions:

- The formality and informality of the scientific communication process;

- The collaborative nature of scientific activities;
- The degree of electronic infrastructuring and advanced networking services.

The regularities encountered in the usage patterns of these technologies might reinforce a structural change in the traditional collaboration practices and communication behaviour in new electronic environments.

2.2.3 Electronic Scientific Communication Reproduces Traditional Structures

The change to electronic networking offers the potential for new and innovative ways of organisation, as well as extended forms of organisation of collaboration. However, empirical evidence is needed in terms of measuring the extent to which traditional or innovative activities are being reproduced or created in electronic forms of research collaboration.

Hypothesis Two: *Electronic scientific communication reproduces traditional structures of scientific communication and research collaboration.*

In order to verify the above hypothesis, as well as to measure the degree of this phenomenon, some of the most typical forms of communication and collaboration should be empirically validated (in non-electronic as well as electronic settings).

First, traditional features of the scientific system (such as the 'Matthew effect' and 'Lotka's law', the hierarchical structure of organisation and specialisation) typical of non-electronic scientific communication and collaboration, should persist within the new digital infrastructures. This being so, we should be able to identify such regularities as characterising the electronic scientific system. However, some indicators might need significant adjustment to cope with the new digital infrastructure.

Secondly, patterns of connectivity in research collaboration, such as significant diversity in the centrality and prestige of key researchers and research groups/institutions, or the significant *clustering* of research communities (identified in networks of citation, co-authorship networks of collaboration or networks of collaborative research projects) should be replicated in the new digital environments. The extent of this structural transformation is also likely to vary across levels of aggregation of collaborative activity (e.g. diversity in patterns of individual researchers connectivity and patterns of institutional connectivity).

Thirdly, the structural evolution of electronic networking should also replicate characteristic processes of "division of labour" and "inter-sectoral" and "international" collaboration known to hold in traditional scientific communities and typical of the development of fields of research.

Finally, science policy initiatives, particularly those focusing both on strengthening the connectivity among researchers and research groups (e.g. networking efforts) and strengthening the electronic connectivity among research resources, should have a structural impact and, therefore, should influence the transformation of traditional structures of collaboration into electronic infrastructures.

Are these regularities of research collaboration, found in traditional non-electronic structures of communication, evolving to new electronic forms and subsequently restructuring features of the traditional science system?

The extent to which new organisational practices need to be developed to allow for the development of the new digital infrastructures for scientific communication is still not known. In fact, we may be going down the wrong path in attempting to identify well recognised organisational practices; significantly different structures of communication might be emerging from electronic interaction.

2.2.4 "Digital Knowledge Bases" are Identifiable Based on the Electronic Reproduction of Research Collaboration Structures

Assuming that research communities are self-organising their patterns of collaboration to form web communities, then we are likely to find important knowledge resources being disseminated and preserved in large-scale electronic networks, such as the Internet or subcomponents of it. This might only be detectable in the medium to long-term, when enduring collaborative relationships among researchers and research groups produce their research results in digital forms made available for wide network use.

Hypothesis Three: *As non-electronic structures of research collaboration are reproduced in electronic environments, we should be able to discover and identify "Digital Knowledge Bases" - large-scale and distributed electronic knowledge structures.*

There are certain conditions that need to be met in order to validate the above hypothesis.
First, collaboration groups, tightly bounded and highly connected, should reproduce their connectivity patterns in electronic networks (see hypothesis in section 2.2.3). This

guarantees local connectivity structures and the "self-organisation" of communities within large-scale and whole electronic networks.

Secondly, knowledge resources - in various forms, e.g. scientific publications, research project reports, human resource information as well as artefacts produced by electronic inter-change of information - should be openly available, widely disseminated and utilised in public, large-scale electronic networks. This is more likely to be achieved within academic scientific communities, but is much less certain when norms, practices and reward structures involve strategies of secrecy or private holding of research results. Such might be the case in application-oriented fields of research (and also perhaps in computer speech and language).

Thirdly, specific science policy initiatives strengthening the public availability of knowledge resources resulting from collaborative networking efforts, are likely to contribute in this direction.

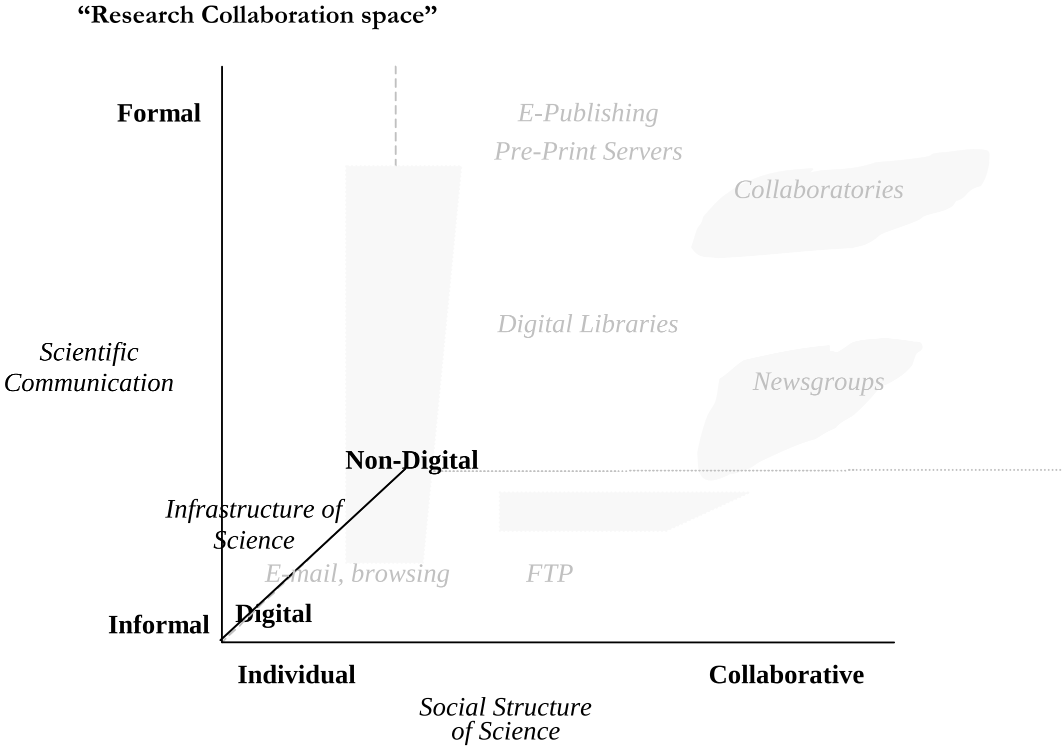
Finally, the production of "digital knowledge bases" does not involve the mere repository of information and knowledge resources disseminated by collaborating entities (researchers, research groups, research institutions). It requires the reproduction of collaboration structures in electronic space, as well as minimum levels of connectivity and boundedness. Nevertheless, important "external" knowledge resources linked by relationships between internal and external entities to the research network, guarantee that these self-organised knowledge zones are not hermetically sealed. Moreover, they provide a mechanism for the expansion of digital knowledge bases into larger systems.

Personal patterns of use of Internet technologies for purposes of research communication and collaboration are likely to be revealed using survey techniques. However, the mapping of knowledge structures on Internet space as well as the diverse "forms" of collaboration structures in electronic environments are likely to demand the use of innovative methodological strategies, based on cybermetric methods in combination with the more traditional bibliometric, network analysis and survey methods. A thorough discussion of this methodological strategy is provided in chapter 3. But, first, in the next section we suggest a new conceptual model for assessment of the structure of electronic scientific communication and its inherent dynamics, integrating the concepts discussed in previous sections and summarising the conceptual framework guiding the empirical investigations.

2.3 A Conceptual Model for Electronic Scientific Communication

The validation of the above theoretical hypotheses would suggest that the transformation of changing patterns of communication and collaboration among researchers in the move to electronic networks and environments is likely to be evolutionary rather than revolutionary. If the structure of electronic scientific communication reproduces existing structures of collaboration, an integrated framework combining the full spectrum of scientific communication activities, scientific collaboration forms and scientific information infrastructures might be represented by the conceptual model depicted in Figure 2.2.

Figure 2-2 - "Electronic Research Collaboration" model



There are some crucial elements in this conceptual model. First, it takes account of the multi-dimensional nature of the electronic research collaboration process. There is one axis or dimension representing the variegated "forms" of research collaboration at all its various levels, from the individual research endeavour at one extreme, to the very complex collaborative research activity, at the other. This dimension is labelled "social structure of science", as it represents the organisation of science and the structure of collaboration activities. There is a second complementary dimension, representing the spectrum of the nature of communication and collaboration activities, from the completely "informal" processes of communication to the formal communication artefacts and activities, such as the end-products of scientific communication in the form of scientific articles or institutionalised research network projects. As discussed above, particular forms of research collaboration, such as "research networks" might be represented as diverse entities along this two-dimensional spectrum - and not as a precise point in a single dimension. Secondly, the model embeds the electronic networking of research collaboration within the structure of the process of scientific communication and collaboration. A third dimension encompassing the digital infrastructure of science is added to the initial "research collaboration space". Indeed, the various "forms" of electronic networking, from inter-personal communication tools such as e-mail, to highly collaborative ones, such as "newsgroups" or "collaboratories", are embedded in the research collaboration space. Here, again, the various forms of electronic networking of research are represented not as single points in space, but otherwise, as volumetric "forms" along a more or less wide spectrum. Thirdly, the conceptual model provides a conceptual representation of the internal structure or patterns of electronic scientific communication. It copes with the multi-modal nature of electronic networking technologies, with the reproduction of traditional collaboration structures into new electronic infrastructures and with the localisation of large-scale digital knowledge bases. Finally, the model can be used for assessing the dynamics of the structure of electronic scientific communication. The evolution in patterns of use of the diversified set of technologies, as well as in the patterns of collaboration and communication, can be represented as an extension to the initial electronic collaboration "forms". We can graphically depict these transformations as different forms in the same model, or comparatively across differences in two or more models.

Sections 5.1 to 5.3 in chapter 5 analyse the empirical evidence testing the robustness and validity of this conceptual model and the underlying theoretical hypotheses. Empirical data are discussed focusing on the individual use by scientists of Internet technologies for research work, electronic collaboration in interactive environments such as newsgroups, and the analysis of knowledge exchange and dissemination of information on the Internet, giving rise to the formation of web communities and digital knowledge bases. In chapter 3 we discuss the methodological strategies implemented in the course of the investigation in order to generate such empirical evidence and to test the theoretical hypotheses discussed in section 2.2. The empirical results are discussed in chapters 4 and 5 of this thesis.

3 Research Design and Methodology

3.1 General Overview of the Methodology

This chapter discusses the methodology adopted in the investigation - the methods and techniques used to gather and analyse the empirical data - as well as the justification for their implementation in order to test the theoretical hypotheses and conceptual model proposed in previous chapters.

The empirical evidence comes from analysis of patterns of communication and collaboration (non-electronic and electronic) in the research field of computational speech and language. This research field focuses on the investigation of how humans and machines might be able to communicate using natural language. Research and development activities include the coding, recognition, interpretation, translation and generation of language. This is a multi-disciplinary enterprise, requiring expertise in the areas of linguistics, psychology, engineering and computer science. The social impact of developments in this research field is tremendous. In fact, computer speech and language (or human language technologies) play a key role in the information age. As referred to in a survey of the state of the art in these technologies (Cole *et al.*, 1997:ii) "the benefits of information and services on computer networks are unavailable to those without access to computers or the skills to use them. ... As the importance of interactive networks increases, those who do not have access to computers or the skills to use them are further handicapped from becoming productive members in society".

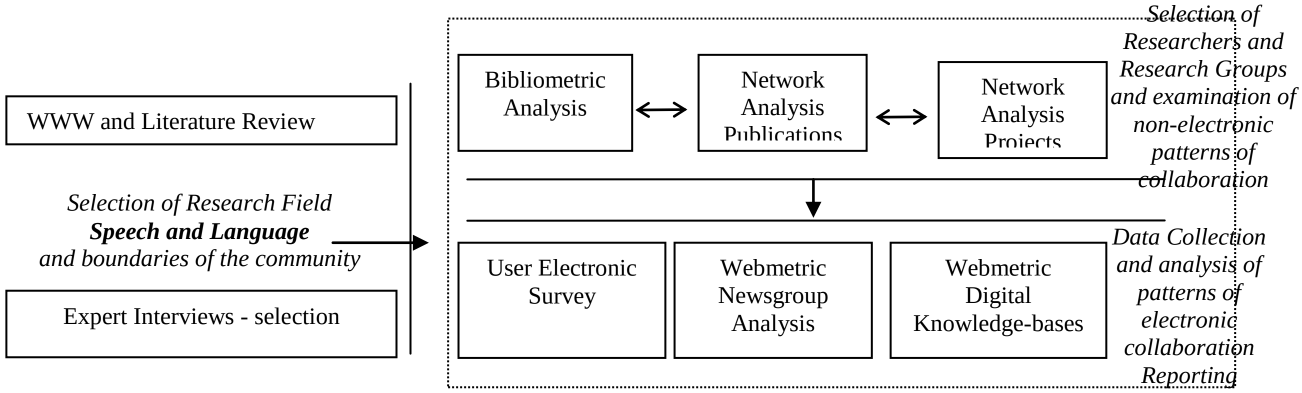
Several complementary factors justify investigation into this specific research field as being particularly well aligned to our analysis of improvements in research communication and collaboration arising from the use of digital infrastructures. This community is likely to use ICTs and the Internet intensively for research work and researchers active in the community are more likely to reliably characterise their use of ICTs for scientific communication and collaboration. This research field constitutes a highly dynamic field, with researchers working at the knowledge frontier, with a very intensive pattern of scientific communication, revealing permanent changes in the structurring and restructuring of research groups and sub-communities. Apparent within the community is the "networking" of research actors - researchers, research groups and research institutions - and research resources - information, specific applications, remote databases, etc. As noted above, the computational speech and language constitutes an inter-disciplinary research field with intensive inter-sectoral and international collaboration. Moreover, we can using relatively efficient methods, identify a European "boundary" to the research field, which facilitates the delimitation of the empirical analysis. The network form of organisation at the European level (and corresponding science policy initiatives), is particularly suited to the analysis of particular forms of organisation and particular patterns of collaboration within this scientific community.

The empirical research of this thesis mainly focuses on analysis of the European Speech and Language research network. The delimitation of the boundaries of the European research network is given by a combination of the analysis of an "institutionalised" research network funded by the European Commission (ELSnet network - European Network of Excellence in Speech and Language Technologies), comprising 141 research centres across 27 countries in Europe, with bibliometric data coming from a co-authorship analysis of the ISI - SCI database (for the period 1981 - 2001). This institutionalised research network has received significant research funding from the European Commission (as revealed by the long-term analysis of research contracts). The structural patterns of collaboration, as given by research project partnerships, are also analysed over the period 1990 - 2002. The empirical analysis of the ELSnet network is significant for various other reasons: first, at the European level, it covers a wide spectrum of the researchers, research groups and institutions dedicated to research in the topic of computer speech and language. An estimate of around 90% - 95% of the whole community is fairly rigorous if one considers that from the biliometric co-authorship analysis less than 5% of the set of identified non-private european groups were not ELSnet members; secondly, it is indeed a Network of Excellence, whose origins date back to 1994 and, as such, the collaboration effects derived from its functioning are likely to be manifest at this time; thirdly, it is an inter-disciplinary and inter-sectoral environment for research communication and collaboration; fourthly, given the absence of a research directory listing active researchers in this field, and given the application-oriented nature of the field of research - which sets obstacles to using only bibliometric approaches to delimit the research field - the Network, in itself, constitutes a roster of research groups and consequently of researchers active in this topic; finally, the research network provides a rationale to link the analysis of patterns of collaboration at the level of the individual researcher, with patterns of collaboration and communication at the level of research groups and institutions.

A key problem in understanding communication and collaboration in these digital environments, and particularly electronic research collaboration, is that the current set of methods for understanding collaboration are not fully able to account for new forms of electronic interaction. New methods for exploring patterns of communication in electronic environments need to be developed. This investigation brings together a range of traditional methods for understanding scientific communication (bibliometric methods, network analysis methods of publications and research contracts and an electronic survey) with an innovative set of methods and techniques for understanding electronic communication and collaboration (cybermetric methods for analysing electronic networks and newsgroup electronic interactive environments).

The following diagram (Figure 3-1) summarises the methodology.

Figure 3-1 Synthesis of the Methodology



As summarised in the opening part of this section and presented above in Figure 3-1, the research methodology is based on a combination of several complementary research methods:

- **Expert Interviews** with two researchers selected for their expert knowledge of this research field and their role as "networkers" or individuals involved in science policy networking efforts and their accessibility, (University of Sussex - a member of Executive board in the ELSnet Network, and University of Brighton - a member of Euromap - both involved in projects engaged in dissemination and networking activities within this scientific community). These experts were expected to provide good information about the development of their field of research and the composition of the scientific community, namely: information about a potential directory listing of active researchers in the field or institutionalised research networks; the importance of formal or informal means of scientific communication, such as the most important scientific journals and conferences; the current stage of development in the field and most important challenges, as well as the diversty of topics being pursued (e.g the extent of specialisation).
- **Bibliometric analysis** of scientific publication in the fields of "Speech and Language" research. The data were collected from the Science Citation Index (1981

- 2001). The aim, as indicated in Figure 3-1 was twofold: firstly, to identify and delimit the network of researchers in this field; and secondly, to characterise particular forms of organisation of the communication process and patterns of collaboration in this field. The search strategy involved a combination of “Keyword search”, “Journal selection” and Co-citation analysis of the most cited authors. The results provide a longitudinal analysis of the formation and development of the field of research, as well as a structural mapping of co-authorship collaboration groups. These bibliometric collaboration groups allowed the identification of researchers and research groups for more detailed analysis using complementary methods (network analysis and a survey of active researchers).

- The Bibliometric analysis was supplemented by a thorough **Network Analysis** of the previously identified clusters, in order to generate indicators of the *connectivity*[20] of research collaboration structures, and identification of *cohesive subgroups*[21] and the *centrality*[22] of certain sub-groups and researchers within the larger networks. (Persson, 1995; Melin, 1996; Newman, 2001).
- The collaboration structures resulting from research **Funding by the European Community** (in the period 1990 - 2001) were also subject to network analysis in order to characterise patterns of research collaboration, as well as to identify clusters of collaboration that could lead to active research groups and researchers in this field. The analysis of collaboration networks resulting from research contracts was considered of relevance for various reasons: first, the author had on-line and non-expensive access to significant amounts of data on research funding in speech and language, for a reasonable period (about a decade); secondly, as this information is very well organised by the European Commission, it is likely to represent the total european funding to the field of research; thirdly, whilst the research funding from the European Commission is likely to represent only a percentange of the total research funding to the field of research, the structural impact to the community is likely to be highly significant (taking into account results from previous analyses, *e.g.* Widhalm *et al.*, 2001); fourthly, the ELSnet members are very well represented in the activities conducted under european

[20] *Connectivity* is a network analysis indicator of the proportion of existing links compared with the total potential links, at different levels - among actors, groups of actors or across the whole network.
[21] *Cohesive subgroups* within a given network are groups of actors that reveal a high degree of connectivity between themselves, and whose relationships with other external groups are weak.
[22] *Centrality* is a network analysis indicator of how central a given actor or group of actors is within the larger network. This is commonly assessed by complementary measures: the *degree of centrality* or activity of actors/groups given by total number of links; the “*between centrality”* or number of paths the actor/group interlinks; and *prestige*, or number of links received by an actor/group, from other actors in the network.

funding. There was at least 1 ELSnet member in 94% (143) of the research projects specifically addressing speech and language (total of 152 projects). This group includes the research programmes in Framework Programme 3 and before that (in a period prior to 1994), the Language Engineering Programme (1994 - 1998), and the Human Language Technologies Programme (1998 - 2002). And at least one ELSnet group participates in 80% (46) of the total number of research projects (59 projects) from programmes related to speech and language but not specifically focused and conducted under the ESPRIT programme (1994 - 1998), the Multilingual Information Society Programme (1996 - 1999), the INCO International Co-operation Programme (1994 - 1998) and the E-Content Programme (2001); finally, ELSnet members were found to be very well connected and central research groups in the connectivity maps resulting from research funding, which is relevant for our analysis of patterns of research collaboration and communication.

The information collected from the interviews with expert informants, as well as that obtained from the bibliometric analysis of publication, the co-authorship analysis of bibliometric data, and the analysis of collaborating groups participating in European funded R&D projects, provided a good basis for understanding of the formation and development of the field of computational speech and language. These methods also allowed a good identification of research groups and researchers active in collaborative activities, as well as the characterisation of patterns of research communication and collaboration (the connectivity of collaborating groups, differentiation in terms of prestige and centrality, specialisation of research groups and division of labour, and hierarchical structures of collaboration). Moreover, the previous analyses provided information about an institutionalised European research network in speech and language, involving 141 research groups across Europe. This information provided the rationale to conduct further analyses in the European network, as well as the needed link between research groups and researchers, for surveying active researchers about their non-electronic and electronic collaboration activities.

To analyse electronic research collaboration in detail, the implementation of the survey to active scientists in the field proved of great value and was a critical component of the whole methodology. Nevertheless, individual patterns of ICT use and electronic interaction, identified from analysis of the survey data, needed to be triangulated by detailed examination (through observation and experimentation) of particular environments for electronic collaboration. This investigation focused on two specific electronic infrastructures: first, interactive environments particularly suited to electronic collaboration and informal electronic interaction, given the open and distributed nature of Internet technologies (see Newsgroup analysis); and, secondly, the assessment of Internet environments (particularly the World Wide Web) as an infrastructure for the dissemination of information, distributed interaction and support of patterns of collaboration, some of them originating from non-electronic interactions (see the section on digital knowledge bases).

- An **Electronic survey** was administered to ELSnet members. This network comprises 141 research groups across 27 countries in Europe and is estimated to represent about 90% - 95% of the total number of european research groups in this field. The electronic survey was undertaken to provide information that could not be directly obtained through the previous two methods (bibliometric and network analysis) and involved a complete census of scientists from the ELSnet network (with a response rate of 21.7% of the total population of 1 431 researchers). The survey was organised into three complementary sections providing information at the level of researcher about non-electronic and electronic patterns of collaboration: (1) disciplinary background of researchers and self-reporting of scientific productivity; (2) interdisciplinary nature of the collaboration as reported by researchers; and (3) ICT usage patterns for scientific communication, scientific work and collaboration. The results emphasise the importance of studies on the use of ICTs by scientists as a mechanism for a better understanding of the process of electronic scientific communication.
- **Webmetric/Cybermetric analyses** of electronic network usage by scientific and technical communities, building upon novel techniques for the generation of quantitative and qualitative indicators of digital infrastructures usage. Among other techniques this includes: (1) data-mining of Newsgroups and electronic discussion fora messages in speech science and technology; and (2) mapping of digital networks based on automatic examination of web links between research centre web sites belonging to the ELSnet network (mapping of electronic connectivity of the total set of 141 web sites).

A principal motivation for employing the methods described above was to develop a structural analysis, i.e. an investigation of regularities in the relationships among entities (be they co-authors of papers, mutual participants in research projects, contributors to electronic environments, or whatever). Specific social network analysis methods were widely used, namely techniques for detecting cohesive subgroups of entities, techniques for measuring the centrality of entities within networks as well as the connectivity of individual entities or of the whole research network. Structural analysis was considered to be the only suitable approach for assessment of the reproduction (or non-reproduction) of traditional patterns of communication and collaboration in electronic environments. The structural approach, focusing on analysing patterns of interaction among researchers and research groups, was expected to provide empirical data of this "reproduction effect", and consequently good empirical evidence for testing the two theoretical hypotheses (Hypothesis 2 on the reproduction of traditional patterns of interaction in electronic environments, and Hypothesis 3 on the dissemination of information revealing large-scale and distributed electronic patterns of collaboration). The empirical data resulting from the implementation of the web-based survey of active speech and language researchers was expected to provide detailed information about patterns of use of Internet technologies for the wide spectrum of formal and informal activities, as well as individual and collaborative research activities. (This conceptual framework was described in section 2.2)

A more detailed examination of each of the research methods is provided in the following sections of this methodology chapter. In section 3.2 bibliometric and network analysis methods are discussed, based on an examination of publications in the field of speech and language. Section 3.3 analyses the collaboration structures of the European speech and language community, focusing on network analysis methods applied to research funded projects. Section 3.4 discusses the methods for assessing patterns of use of Internet technologies in the process of scientific research and collaboration. Finally, in section 3.5 cybermetric methods developed to generate web indicators of scientific research are examined. These methods are applied to empirical data on electronic research networks.

3.2 Bibliometric and Network Analysis of Speech and Language Publication

3.2.1 Publication in Scientific Journals

In order to identify and then to analyse the structure of formal scientific communication as well as to identify patterns of collaboration and their evolution over time, a bibliometric analysis of publication in scientific journals was conducted for the field of computational speech and language. This analysis involved two complementary stages: the first involved a search strategy to identify a subset of relevant publications from a bibliographic database, while the second took the form of a co-authorship network analysis of patterns of collaboration as revealed by the formal communication process that co-authorship represents.

The bibliometric analysis consisted of an analysis of publications as referenced by the Institute for Scientific Information, Science Citation Index database, in the electronic version available from the Web of Science at the following URL http://www.mimas.ac.uk/. Publications cited in this database were searched and downloaded in the period from January to March 2001.[23] The publication records were collected for the period 1980 - 2001 using a combination of several search strategies. Given the interdisciplinary and application-oriented nature of this field of research, the use of citation classification systems or a single search strategy was found to be inadequate. Consequently, the broad search strategy for defining the base bibliometric set combined the techniques of *Keyword Search*, *Core Journal* publications and the publications of *Core Authors* (this last applied to speech research but not to language research, given that background knowledge about specific contributions, collected from the expert interviews, was only available for speech and not for language). The following table summarises the bibliometric strategy.

Table 3-2-I - Summary of Bibliometric Extended Broad Search Strategy

Keyword Search
A limited set of keywords was defined for searching the citation database, for every year from 1981 through to 2001. These keyword sets were defined after consultation with two experts, as well as after analysis of bibliographic sources surveying the state of the art in speech and language research (Cole *et al.,* 1997). The following sets were selected, for publications including those keywords in the *Title*, *Abstract* or *Keyword* fields: For Speech research: (Speech AND recognition) or (Voice AND Recognition) or (Speech AND Synthesis). For Language research (Language AND Processing) or (Language AND Generation) or (Language AND Understanding).

[23] In the period September to December 2000, a first pilot bibliometric study was conducted in the sub-field of Speech Recognition. This publication data set was from the BIDS service, available on-line at the following URL http://www.mimas.ac.uk/bids. However, as the BIDS service was superseded by the Web of Science, the full and enlarged bibliographic data set was collected for the later period.

Core Journals Search
After consultation with experts as well as bibliographic analysis, two sets of journals were selected for retrieving publications for speech research and for language research respectively. Within each research field a distinction was made between general journals and specialised journals. The following journals were analysed. For Speech Research: Specialised - *Speech Communication* and *IEEE Transactions on Speech and Audio Processing*; General - *Journal of the Acoustical Society of America* For Language Research: Specialised - *Computational Linguistics*; Generalised - *Journal of Artificial Intelligence Research* and *Communications of the ACM*. All the publications from *Specialised* journals were selected for the bibliometric data set. General journals were searched for publications containing the above keyword sets.
Core Authors Search
Given the existing background knowledge for the sub-field of speech recognition, an additional set of publications was added to the Speech research bibliometric data set. Publications related to three seminal contributors to this subspecialty were considered: the co-authored core of Rabiner; the citation core of Rabiner; and the related core of Rabiner, Bahl and Juang (highly cited authors in Automatic Speech Recognition during the period 1981 - 2001). The co-authored core is defined as all the publications authored by Rabiner, as well as publications authored by direct co-authors. The citation core of Rabiner consisted of all the publications citing Rabiner's seminal article in 1989. The related core of the three authors consists of the set of publications which have one or more identical references with the references cited in an author's "seminal article" - the criterion for this being the most cited article - of the core author. This allowed the selection of "bibliographically-coupled documents" with *seminal articles* of core authors. For language research, no set was defined, as in this subspeciality no significant "revolutionary" contribution was identified by the "expert informants" as having such a tremendous impact on the development of the subspeciality, as occurred in speech recognition with Rabiner's work.

The above search strategy allowed the delimitation of the bibliometric data set for speech and language research, subject to co-authorship analysis.

3.2.2 Network Analysis of Co-authorship

A network analysis of co-authorship data was conducted. There were two main goals of this analysis. The first was the characterisation of collaboration patterns and collaboration groups based on the formal system of scientific publication. Co-authorship is taken as a partial measure of collaborative relationships within a community. The extent to which co-authorship accurately reflects collaborative links cannot be ascertained without other corroborating information such as might be gained by a survey or from Webmetrics - it may be the case, however, that co-authorship relationships *do* accurately reflect the structure of the network despite the possibility of other relationships and in this case the other relationships may be seen as complementing or augmenting the co-authorship relationship. The evolution of these collaborative structures over time was also considered in the analysis. The long-term analysis of the evolution of co-authorship structures provided good empirical evidence of the development of each subspeciality (speech and language). Equally important was the differentiation among these collaborating groups of more centrally located and better connected research groups within these networks. A second goal of the co-authorship analysis was the identification of collaborating groups belonging to the European speech and language research area, as well as potential collaboration groups considered as important "external" world authorities in the field.

The network analysis involved the detection of cohesive subgroups of authors, who collaborated more intensively. "Cliques" were identified for each research area over several periods. In addition, the co-authorship networks for different periods were analysed in terms of the centrality of the subgroups (sub-groups with more activity within the group and with other subgroups) as well as the best connected collaborating groups. (A "clique" is a tightly bounded subgroup of the co-authorship network, whose members collaborate at least once with each of the other members of the group.)

In depth network analysis of the bibliometric data allowed the identification of several important research groups within the European area, as well as important "reference" international teams in speech and language research. However, as the formal communication system reveals only a subset of the full spectrum of research communication and collaboration activities, a network analysis of the collaboration patterns resulting from European funding was carried out.

3.3 The Collaboration Structure of the European Network for Speech and Language Research

The bibliometric analysis and corresponding co-authorship analysis identified groups collaborating in computational speech and language research within the European area. The methods discussed below (network analysis of European funding and the electronic survey of the ELSnet network) were applied to the European area only (i.e. focusing on European research groups). The network analysis of European funding in Human Language Technologies[24] over the period 1990 - 2001 complemented the previous analysis of bibliometric data.

The fundamental purpose of analysing the structures of collaboration resulting from research projects funded by the European Commission was to examine the characteristics

[24] In this investigation, the concepts of "Human Language Technologies" research and "speech and language" research are used interchangeably. A finer distinction is provided in Cole *et al.*, 1997.

of connectivity among different research groups, as well as the structural evolution of those connectivity patterns over time. Connectivity in this regard is meant as the extent of linkages among research groups, at a certain point in time as well as over time, in terms of the frequency with which those research groups collaborate with each other. Potentially, the differences among collaborating groups within the network in terms of their centrality and connectivity are likely to reveal the essential characteristics of the process of scientific communication and collaboration (such as differences in prestige, or more or less institutionalised hierarchies within the scientific community). However, these connectivity differences might be related to the processes of division of labour and specialisation within the research community.

Information on the entire set of research projects funded by the European Commission in Speech and Language was available electronically at the following URL http://www.hltcentral.org/projects. These research projects cover the Third, Fourth and Fifth Framework Programmes of the European Commission. The data set comprising 211 projects and involving 1,071 different research institutions, was downloaded during the period March to June 2001. The analysis provided empirical evidence on European investment in this research field, as well as on the persistence of certain institutions as central members of these European research networks. The institutional patterns of connectivity are likely to reveal the differentiation of research groups within the research area, in terms of both the activity and more intensive participation in research collaboration activities, as well as in connecting collaboration activities with other network members. Each of the seven research programmes analysed has given rise to a separate research network. Network analysis of each network was conducted in order to determine the subgroup structure of the whole network (identifying and characterising the subgroups within the whole network), as well as the centrality (in terms of activity and connectivity) of subgroups within the network. As the longitudinal data set allowed for analysis of the evolution of these structures of collaboration, and identification of persistent patterns of connectivity and centrality of some research institutions.

The patterns of institutional connectivity and research collaboration (as given by research funded projects) were then compared with the characteristics of collaboration as given by individual scientists in an electronic survey of the researcher members of the research groups participating in the ELSnet network. A brief discussion of the survey method is provided in the next section.

3.4 Survey Analysis of "ELSnet Network". Collaboration Patterns and Use of Internet Technologies

An electronic survey was administered to researchers who were members of research groups participating in the ELSnet network - the European Network of Excellence in Speech and Language Research. This network comprises both a research network funded within the European area and an extensive roster of the whole population of researchers in this topic. Moreover, this "roster" was tightly delimited by the institutional affiliation to the network, which was a fundamental reason for its adoption as the survey population.

The ELSnet network connects 85 academic and public research centres and 56 private research laboratories, representing, respectively, 60.3% and 39.7% of the total number of research centres. A total population of 1,920 researchers was identified through analysis of the research centre web sites. However, of these 1,920 researchers only 3.4% belonged to private laboratories. The main reason for this is that private research centres do not give out contact details for their research staff. The survey population is thus largely confined to the academic research community (but including public laboratory researchers). For only 1,431 researchers was it possible to identify a personal e-mail address at the time of the survey. The survey was administered between June and September 2001. In order to guarantee a better response rate to the survey institutional endorsement from the ELSnet coordinating group was sought and granted. The first e-mail message, sent on the 27 July 2001, invited respondents to participate in the survey. Four additional recall messages were sent, the final one being dated up to 25 September 2001.

The invitations pointed to the electronic questionnaire available at http://www.sussex.ac.uk/Users/prpb7/speech/survey/ELSnetsurvey.htm. A full copy of the electronic questionnaire is available in Annex X. The electronic survey was implemented in a web form which the researchers could fill in and through a form handler application, despatch the data directly to the e-mail address of the author of this investigation. Access to technologies, other than access to the web using a normal web browser, was not necessary. The selection of a well managed e-mail system to receive the results from the web-based survey was justified by experience of secondary problems related to sending files attached to e-mail messages and to avoid security threats to the computers of potential respondents. Moreover, this technique guarantees a reasonable control and monitoring of the pace of the survey and response time.

The survey was structured in three parts. Section I focused on background information from the researchers (such as institutional affiliation and disciplinary background as defined by post-graduate specialisation, as well as a self-assessment of the research speciality and research area in which respondents carry out their research). Section II focused on research activity and performance as well as individual patterns of collaboration, including a quantitative record of publications and research projects and the identification of collaboration patterns and the inter-disciplinary nature of their collaboration work. Section III analysed personal communication behaviour with regard to usage of Internet technologies for research work (the range of technologies used, comparison of traditional communication technologies with new Internet technologies, intensity and diversity of technology use for different stages of the research work, individual and research group assessment of Web use as a repository for research information and a platform for research collabortion, type and characteristics of the use of more advanced technologies, such as remote servers and high-performance computational tools). A total of 312 completed questionnaires (representing a response rate of 21.7%) was analysed. The extensive data set provided good quality data (in terms of representativeness of the sample - both geographical and in terms of seniority of active reseachers) for all three sections described above.

The combination of biliometric analysis with co-authorship analysis and network analysis of research collaboration projects within the European area, as well as with the individual survey information gathered from ELSnet researchers, provided good empirical evidence on scientific communication behaviour and the structural patterns of research collaboration and connectivity of researchers, research groups and research institutions.

A final set of cybermetric methods was used to "triangulate" the information obtained from the survey with regard to usage of information technologies by scientists for research collaboration and scientific communication, in order to assess the accuracy of the hypothesis concerning the similarity of structures of communication between non-electronic or traditional and electronic scientific communication.

3.5 Innovative Cybermetric Methods

A fundamental assumption of this investigation was that a detailed analysis of patterns of electronic networks usage by scientists required the collection of empirical evidence based on innovative cybermetric[25] methods. In fact, while traditional survey methods (see section 3.4 above) allow one to assess the personal communication behaviour of Speech and Language researchers, a complementary assessment of patterns of electronic network usage for collaboration and electronic connectivity demands analysis of longitudinal data sets resulting from practice within electronic environments. To facilitate this, the investigation focused on the analysis of two such environments: interactive electronic discussion fora (widely known as "Newsgroups") and the electronic connectivity of research centre web sites (Web structure analysis, restricted to the ELSnet network). The reasons for focusing on the analysis of these electronic environments were as follows. First, the combination of the two methods allows the structure of electronic communication to be mapped along a broader spectrum of the previously identified dimensions of formality/informality and individual/collaborative processes of communication and collaboration; secondly, both methods are likely to be widely used by the research community; thirdly, patterns of informal scientific collaboration and communication can be assessed by analysing newsgroups over a reasonably extensive period; fourthly, the structure of electronic connectivity of research centres (as given by web linkages) can be compared to the structures of non-electronic research collaboration; and finally, the empirical evidence is freely and openly available for reasonably long periods, and is non-intrusive if used for statistical purposes. Each of the cybermetric methods is described below.

3.5.1 The Structure of Interactive Electronic Communication: Newsgroups

The data-mining of newsgroup discussions was based on the analysis of 28,185 electronic messages exchanged in three specialised fora of speech science and technology (*comp.speech*, *comp.speech.research*, and *comp.speech.users*). An archive of newsgroups messages for the three specialised newsgroup fora (comp.speech - from 1992 until 1998 and comp.speech.research and comp.speech.users, for the period 1998 - 2000) was recovered electronically from ftp://svr-ftp.eng.cam.ac.uk/pub/comp.speech/archive/. Using a Visual Basic application

[25] In this investigation, the term "cybermetric" is used interchangeably with the term "webmetric". However a stricter definition of "webmetric" methods would restrict such methods to analysis of patterns of use of the "Web" or "World Wide Web", whereas the more global concept "cybermetric" would entail measurement of the whole set of Internet technologies and services (e.g. *E-mail* use, *Newsgroups* use, *File Transfer Protocol* use, *Video-Conferen*ce use).

developed by the author,[26] these messages were filtered to generate variables including message ID, message type (original discussion or reply message), author's name, author's e-mail, author's organisation, date (year, month, day), subject, and text of the message. Some additional variables were calculated from these, namely the country and region of origin of authors, based on the author's e-mail and the author's organisation, and the type of organisation (industry, university, other) based on the organisation field and e-mail of the author.

The analysis of this large data set of newsgroup messages focused on three complementary dimensions. First, various *quantitative and qualitative indicators* were computed for the participants of newsgroups. Among these were: the *size* of these newsgroup systems, as given by number of messages, number of discussions, number of contributors, etc.; *time-series* variables that allowed the longitudinal analysis of newsgroups and comparison with typical diffusion processes; *activity and prominence* of the participants, based on the number and type of messages, allowing distinction to be made between contributors to these electronic systems; and the analysis of *inter-sectoral and inter-national collaboration* as given by the type of organisation and nationality of the contributors to the newsgroup's discussions. This first set of newsgroup indicators allowed the identification of important members of the research community, participating and "leading" newsgroup discussions. A particular analysis of the most active discussions, and the most active researchers in the newsgroup fora, allowed the identification of ELSnet researchers (5 ELSnet researchers in the set of most active 39 contributors with more than 10 participations), some of them directing ELSnet research groups (IBM France, University of Geneva, FORWISS in Germany, Katholic University of Sweden, Technical University of Berlin). Naturally, the public and open nature of these systems, as well as the focus on speech science and technology historically leaded by US research groups, explains why such newsgroup activity is not restricted to european researchers, but nevertheless it was good to find members of the ELSnet community among the well connected members of these electronic systems.

Secondly, a *network analysis* of the three newsgroups was carried out that allowed the mapping of the structural evolution of these systems. This analysis allowed the characterisation of groups and social structures emerging from frequent interaction within these electronic environments. Moreover, the structural evolution of these electronic

[26] Details of the Visual Basic application for cybermetric analysis of Newsgroups are available from the author, and temporarily available at the following URL http://www.sussex.ac.uk/Users/prpb7/DigitalTools/NewsNetworks.htm.

systems was assessed and the process of "specialisation" and "division of labour" between subgroups was analysed.

Thirdly, the *content analysis* of newsgroups messages facilitated the identification of the content being exchanged in these electronic interactions. Extensive analysis of several random samples of newsgroup messages allowed the determination of an embryonic typology for the content exchanged within these electronic environments.

The empirical analysis and discussion of the newsgroups methods and corresponding results are provided in chapter 5, section 5.2, where the more fundamental theoretical hypotheses concerning the evolution of newsgroup systems, the structure of electronic collaboration and communication, and the formation of "electronic invisible colleges" are analysed.

3.5.2 Mapping Electronic Connectivity and the Discovery of Digital Knowledge Bases

A second cybermetric method developed in the course of the investigation consisted of a combination of techniques for the structural analysis of the electronic connectivity of research centres. This "electronic connectivity" was measured in terms of the electronic linkages with other electronic resources disseminated on the World Wide Web. The fundamental assumption underlying these techniques is that a research institution to a certain degree reveals its patterns of non-electronic collaboration through its electronic connectivity with other electronic resources (particularly from/to other research centres). This hypothesis was tested by analysing the patterns of electronic connectivity of the whole set of 141 research centres belonging to the ELSnet network.

A combination of powerful search engine techniques, based on Google technology (available at http://www.google.com/) and the development by the author of a web crawler for mapping internet webspace around an initial research centre (Galilei technology,[27] provided the computational tools for collecting the raw data needed for the electronic connectivity analysis. The network analysis of these data sets allowed the characterisation of the structural patterns of connectivity, namely the identification of electronically "best-connected" and "most central" research centres, the identification of

[27] Galilei is one component of a software agents system for mapping electronic networks, allowing the collection of the whole set of links (URL resources) connected to an initial entity web site URL. More information is available from the author, and temporarily from the URL http://www.susssex.ac.uk/Users/prpb7/DigitalTools/Galilei.htm.

linkages of important external electronic resources to the network and, ultimately, the exploration of the concept of "digital knowledge bases". These electronic connectivity analyses are discussed in detail in section 5.3. The detailed discussion of the concept of "digital knowledge bases" is provided in section 5.3.3.

The process of mapping digital networks based on the electronic connections of research centres allows one to identify the electronic connectivity of those research groups at the micro, meso and macro levels. At the micro-level, one can focus on a single research group and characterise the patterns of connectivity of that research group, based on the internal/external connections as given by the web links. At the meso level, one can examine or compare two or more research groups and measure the intensity of mutual connectivity. At the macro-level, an overall network of research centres can be examined and "digital knowledge-bases" or overlapping electronic regions identified, based on the whole set of electronic links.

Non-electronic research collaboration networks are likely to have a digital representation. This electronic representation is not necessarily similar to the non-digital social structures of collaboration. The above cybermetric methods were used to test the hypothesis that the essential characteristics of non-electronic collaboration networks are "extended" to the electronic environment of the World Wide Web. This being so, these large-scale electronic networks are organised in a "sub-communities" structure, the mapping of which reveals the sharing of basic knowledge resources and identifies transactions among research entities. In this sense, a "digital knowledge base" does not constitute merely a repository of information and resources from a group of research institutions. Rather, it represents in electronic environments the collaboration patterns of those institutions and the internal connectivity of their research activity. Moreover, esssential "external" links to important resources (from other institutions in the same field of research) are likely to be identified.

In chapter 5, Section 5.3 an indepth discussion of the methods, data and results on mapping digital research networks is provided. The theoretical assumptions that are empirically verified concern: a) the structural similarity of the patterns of electronic connectivity and non-elecronic collaboration structures for the set of research institutions belonging to the ELSnet network; b) differentiation by degrees of centrality and prestige of research institutions, even among the "core" group of "most central" and "best-connected" institutions; and c) the potential discovery of "digital knowledge bases" through detailed mapping of the electronic networks of research institutions. The first two points partially test hypothesis two (section 2.2.2). The results triangulate previous bibliometric and network analyses (chapter 4, sections 4.1 and 4.2), as well as the survey results, in terms of the electronic connectivity and extension into electronic environments of traditional (non-electronic) patterns of collaboration and communication. The third point above links directly to hypothesis three (section 2.2.4), concerning the identification of “digital knowledge bases” – i.e. large-scale and distributed knowledge structures. This directly links with the results of the survey regarding electronic networks as repositories of knowledge and environments for knowledge inter-change (section 5.1.6), as well as with the Newsgroup content analysis (section 5.2.7) of knowledge-exchange within electronic networks.

Having summarised the combination of methods used for this investigation, a first discussion of some methodological limitations is provided in the next section. A more general discussion of methodological limitations is provided in the conclusion to this thesis (chapter 6).

3.6 Conclusions and Limitations of the Methodology

The methodology for this investigation involved a combination of three complementary sets of methods. These methods were implemented in order to collect and analyse empirical evidence on the structure of scientific communication and research collaboration in both environments - i.e., non-electronic traditional scientific communication and electronic scientific communication and research collaboration. The first set of methods consisted of the bibliometric and co-authorship analysis of publication data in the field of computational speech and language, complemented by network analysis of collaborative research projects funded by the European Union. This allowed an extensive and quantitative analysis of the structure of research collaborations emerging from the formal system of scientific communication represented by scientific journal publication as well as structures of research collaboration directly involved in research project activity. Second, a World Wide Web-based survey was conducted that provided an opportunity for virtually all of the members (some did not have e-mail addresses so did not receive an invitation) of the ELSnet network to respond. This technique allowed analysis of the interdisciplinary nature of the research field as given by researcher’s background and individual pattern of collaboration and communication behaviour as well as detailed analysis of individual use of Internet technologies in the course of research work and scientific communication and collaboration. Third, a set of innovative cybermetric methods was developed in order to "triangulate" the empirical evidence to assess the changing patterns and structures of electronic scientific communication. These cybermetric techniques allowed extensive and detailed analysis of Newsgroups as electronic environments for interactive electronic collaboration as well as detailed mapping of the structure of web linkages among research institutions participating in the ELSnet network.

A general limitation of the methodology centred on the necessity of collecting empirical evidence from two different units of observation researchers (for the bibliometric analysis, survey analysis and newsgroup analysis) and research institutions (for the network analysis of research projects and cybermetric analysis of digital knowledge bases). This limitation was outweighed by the broader spectrum of research collaboration patterns made available for analysis. A second limitation was the absence of a tightly bounded, highly connected *roster* of researchers in the field in which the whole set of methods could have been implemented. In the most part, this limitation was counterveilled by taking as the main population for analysis the ELSnet network – researchers and research groups - which is likely to represent about 90% - 95% of the total european research community. However, a complete roster of researchers would have provided an even better setting for empirical testing of the questions discussed in the conceptual framework chapter (see chapter 2).

Specific limitations of the methods implemented (e.g. problems in delimiting the boundaries of the bibliometric data set for an interdisciplinary field of research like speech and language, or the limitations of using survey data provided by individual researchers for assessing electronic collaboration patterns at higher aggregate levels) are throughly discussed in the respective methodology sub-sections of each chapter.

This investigation seeks answers to the research question of whether or not (and to what extent) traditional research patterns of collaboration and features of scientific communication are being reproduced in electronic environments. Chapters 4 and 5 provide an indepth discussion of the empirical evidence for patterns of communication and collaboration - non-electronic and electronic - in the field of computational speech and language.

Chapter 4 focuses on analysing traditional (non-electronic) patterns of research collaboration and identifying features of communiction in the computational speech and language community. Section 4.1 discusses the results for the bibliometric and co-authorship analyses of the formal communication system. Section 4.2 complements these analyses of the structure of non-electronic collaboration by discussing the results for the network analysis of European funded research projects in speech and language.

To complement this, chapter 5 examines the results of the use of Internet technologies for scientific work, research collaboration and scientific communication. Three particular dimensions are thorough analysed (how these electronic networks support individual and research group-work; how these electronic networks more widely support collaboration activities - formal and informal - and reproduce non-electronic patterns of collaboration and communication; and, finally, to what extent these electronic networks suppport widespread dissemination of scientific information and inter-change of data among collaborating entities. The empirical evidence comes from: a) the electronic survey administered to ELSnet researchers (section 5.1); b) the electronic structure of interactive discussion fora, as given by informal information exchange within newsgroups (section 5.2); and c) the electronic structure of Internet connectivity and the discovery of digital knowledge bases, discussed in section 5.3.

4. Data Analysis and Discussion - Scientific Communication and Collaboration

Chapter 4 analyses in detail empirical evidence on scientific communication and research collaboration in the field of computational speech and language. The discussion is conducted in terms of helping to explain the socio-organisational nature of this community. The changing nature of this socio-organisational context is considered of critical importance for understanding the interactive transformations brought about by the new technological capabilities of electronic communication and collaboration. Moreover, detailed knowledge about the organisation of the research community and particular patterns of communication and collaboration among researchers and research groups, contributes directly to answering the main question in this research as to whether or not traditional (non-electronic) forms of organisation are being transformed (or replicated) within electronic networks.

Section 4.1 focuses on particular features of the formal communication system within this research community, to try to get a better understanding of the formation and development of the research field, and the dynamics and evolution of collaboration patterns resulting from co-authorship within the community. It also attempts to identify collaborating groups of researchers whose communication behaviour reveals specific forms of organisation within the community.

In Section 4.2 we examine further empirical evidence on collaboration both at the level of research groups involved in speech and language research funded in the European area, and at the level of individual researchers participating in the European network of speech and language. The delimitation of the boundaries of this community to the European area allows a closer analysis of particular patterns of individual research behaviour in terms of communication and collaboration. Further analyses relate to the characterisation of patterns of connectivity among research groups, forms of organisation in this community in terms of hierarchical structures, specialisation of research groups and division of labour, and differences in intensity of collaborative behaviour across different parts of the community.

Both sections provide a longitudinal analysis of the evolution of collaboration and communication structures in speech and language, revealing some of the "hidden" social structures underlying these collaborative processes. The knowledge derived is advanced by a detailed analysis of the results of the survey to researchers in this field – their self-reporting of collaborative practices (discussed in this chapter, in section 4.2.2) and their patterns of use of ICTs (discussed in section 5.1 of the next chapter).

The following section begins this discussion by examining quantitative and formal measures of scientific communication revealed by publications in scientific speech and language journals.

4.1 Quantitative Measures of Scientific Communication
The Formation and Evolution of the "Speech and Language" Research Communities

4.1.1 Introduction

This section focuses on the analysis of quantitative measures of scientific communication and particularly the formal system of communication in computational speech and language, as reflected by publications in scholarly journals. The discussion is based on an empirical analysis of the research communities of "computational speech and language", including both speech research and technology as well as language research and technology. The research problems being analysed by this community could be defined as 'a broad range of activities with the eventual goal of enabling people to communicate with machines using natural communication skills' (Cole *et al.*, 1997: 3). This is certainly a multi-disciplinary enterprise, involving competencies from diverse fields, such as linguistics, psychology, engineering and computer science. This is empirically evidenced by the diverse backgrounds of the researchers within this community, the diversity of scientific journals and conferences in which community members share their results, and the diversity of research groups specialising in particular subspecialities of the overall speech and language problem.

Creating machines that interact with people gracefully requires a good understanding of acoustics and the symbolic structure of language (the domain of linguistics), as well as knowledge about mechanisms and strategies of communication among people (the natural domain of psychology). Advances in signal processing are needed to produce robust systems (hence the necessary contribution of electrical engineering) as well as advances in computer science, in order to create the architectures and platforms needed to represent and utilise all this knowledge. The intense patterns of collaboration (at various levels) are manifest whether one analyses patterns of co-authoring behaviour, or observes patterns of cooperation in research and development activities (see empirical evidence of this in section 4.2 below). Nevertheless, different paces of development are well recognised in terms of the wide diversity in the activities of computational speech and language research. (For our purposes, we aggregate these activites into two major categories: speech research, on the one hand - the coding, recognition, and synthesis of speech - and language research, on the other hand - including the understanding, interpretation, translation and generation of language.) Speech research is considered to be at a more mature stage in terms of solving the particular problems of speech recognition and engineering, while major developments in language research did not occur until the 1990s.

The empirical evidence from the bibliometric analysis corroborates these different stages of development (and particularly the revolutionary developments in speech research of the late 1980s). The specialisation of publication in different scientific journals also corroborates the differences in organisation of these two complementary subcommunities. Nevertheless, the long-term analysis of publication and the evolution of collaboration patterns (based on co-authorship networks), as well as the "networking" effects of collaboration in cooperative research and development projects (see section 4.2) reveals intense forms of cooperation among speech and language researchers and research groups. A final remark setting the context for understanding the particularities of speech and language research concerns the special position of Europe in terms of the challenges in language research (for the challenge of multi-lingualism in Europe, see e.g. Mariani 2000). While speech research is generally recognised to have been led by work in the U.S., the language diversity in Europe with all its consequent problems and challenges has steered Europe into a strong position to lead language developments. The networking effects of collaboration (partly incentivised by European science policy) are thoroughly analysed in section 4.2.

In this section we start with the bibliometric analysis and then move on to the collaboration patterns established by formal communication. Empirical data were collected from ISI - SCI and analysed from a bibliometric study of "speech and language" (ASR - Automatic Speech Recognition and NLP - Natural Language Processing) and associated co-authorship networks over the period 1981 - 2001.

It is of particular interest here to understand how and to what extent quantitative measures of scientific communication, namely bibliometric data, reveal the structure of research collaboration within an inter-disciplinary, technology-driven and application-oriented field of research. It should be stressed from the outset that the bibliometric techniques are not being used for the purposes of research evaluation. In fact, we are not interested here in the mapping of research productivity based on science maps. Neither are we interested in the construction of "knowledge-maps" of these fields of research. The specific focus is on discovering the underlying collaboration structures that produce and disseminate scientific knowledge, the most visible form of which are scientific journal articles. We chose not to pursue the elaboration of the epistemological structure of the field (e.g. to provide detailed classifications of knowledge types and their relationships) given our research focus on patterns of connectivity and collaboration among researchers and research groups.

The research question could be stated as follows. To what extent can we identify the structure of research collaboration based on analysis of formal scientific communication? And what additional problems are we likely to expect when analysing inter-disciplinary and application-oriented fields of research?

The analysis is conducted by tackling four particular research issues. First, it is necessary to discuss methodological strategies for overcoming the problem imposed by the absence of a rigorous *ex-ante* delimitation of the boundaries of fields of research and research collaboration groups. This problem is heightened when the scientific speciality is inter-disciplinary, technology-driven and application-oriented by nature, as is the case with speech and language research. Secondly, we examine the particular significance and structural impact within the community, of highly important "revolutionary contributions" to the development of a scientific speciality. Thirdly, we consider methodological contributions for mapping the evolution of research collaboration networks using a combination of bibliometic methods and social network analysis. Lastly, we look at the mapping of the structural dynamics of two inter-connected communities - the "speech" community and the "language" community - and the potential identification of research groups and institutions within the European area.

The discussion is organised as follows. In section 4.1.2 we review the literature covering the more general arguments on the formal system of scientific communication and discussions about the reliability of bibliometric methods, and the more particular analysis of co-authorship studies for mapping research collaboration structures. Section 4.1.3 details the research question and operational hypotheses underlying the investigation in this section, as well as the methods used for the collection and analysis of empirical evidence. Section 4.1.4 discusses the limits and bibliometric methodologies for delimiting the boundaries of inter-disciplinary fields of research. Section 4.1.5 tests the hypothesis "The winners 'structure' it all", concerning the structural impact of highly important contributions and prolific authors. Section 4.1.6 discusses the dynamics of research collaboration structures, as reflected in the evolution of co-authorship networks. In section 4.1.7 the results of the co-authorship analyses extracted from the longitudinal analysis of publication, and the potential identification of European research groups on computational speech and language are examined. Finally, in section 4.1.8 the main conclusions and further lines for investigation are briefly summarised.

Some of the results which set a conceptual framework for the rest of the discussion can be predicted. The structural analysis emphasises the importance of combining these advanced bibliometric methods with other research techniques, such as surveys of researchers in the community. In fact, the longitudinal analysis of co-authorship activity is likely to reveal only a portion of the "hidden" structures of collaborative activity. Information gathered directly from active scientists in the field is likely to highlight important features of communicative and collaborative behaviour. Another important factor is the involvement of "experts" in the assessment of "research collaboration maps" as well as during the planning stages of the mapping process. In the absence of such an "expert" assessment, the analysis of other "collaboration forms", such as participation in research funding programmes, and contribution to more "informal" communication fora (e.g. international conferences, symposia and seminars or newsgroups), may also prove to be significant in highly dynamic and evolving fields of research. This is the case with computational speech and language.

4.1.2 The Formal System of Scientific Communication and the Structure of Research Collaboration

In this section we briefly review a branch of the literature on bibliometric methods focused on using co-authorship and network analyses for mapping inter-disciplinary fields of research. An essential problem that remains unsolved is the potentially more rigorous delimitation of the boundaries of "inter-disciplinary" fields of research. Even though it is not strictly based on bibliometric data (most of the analysis is supported by sociometric data collected from a survey), a recent study by Menéndez *et al.* (2001) on the multi-dimensionality of interdisciplinarity is particularly important here. The three complementary dimensions under analysis are the diversity in personal training and research specialisation of scientists, the research practices and collaboration behaviour of the groups, and finally the cognitive inputs and outputs of the research activity. This last dimension was analysed using bibliometric data collected both from the survey and from an analysis of the "core" journals in the three specialities under analysis (cardiology, pharmacology and materials science). The authors stress that the use of bibliometric data for collaboration analyses might be inappropriate if based on the use of arbitrary delimitation of disciplines for classification purposes. In an analysis of cardiovascular biology, Rogers and Anderson (1993) devised a strategy to overcome the problem of defining a multi-disciplinary field of science. They suggest a three-stage process for defining a boundary to the research field. First, they use appropriate sections of a hierarchical classification scheme as a filtering mechanism in selecting publication records from an indexing database. Secondly, a selected set of documents is reviewed by a panel of "experts" to provide a more sophisticated filtering process. Thirdly, the publication lists are validated with lists obtained from academic departments working in the field. The combination of these procedures proved to be a reasonably robust methodology for assessing the boundaries of the publication set, and the consequent identification of the major departments working in this area.

Finally, the last exercise specifically focused on testing the robustness of a new strategy for mapping multi-disciplinary fields of research using co-authorship analysis. Zulueta and Bordons (1999) suggest a non-traditional approach based on a novel searching strategy for delimiting the boundaries of fields of research not fitting within the traditional disciplinary boundaries. The suggested search strategy is based on a combination of a title keywords search, an institutional address search and a selection of journals. This strategy was compared against the more traditional "core journals" selection strategy, and yielded better results, allowing the identification of 40% more research groups working in cardiovascular research in Spain. The results were validated by "experts" in the field of research who confirmed the robustness of the methodology. One interesting finding in this study was that the "broad strategy" was especially useful for the identification of university teams, since they published more often outside the "core" set of journals. An additional result was that the broad strategy had a relatively poor "precision ratio" in terms of including false positives in the whole data set, but, nevertheless, allowed a more comprehensive map of research collaboration in this field.

In the next section we discuss the research question, hypotheses and methods guiding the empirical discussions of sections 4.1.4 to 4.1.7.

4.1.3 What is Revealed by Formal Scientific Communication?

The above literature review reveals some important gaps in the process of identifying collaboration structures (groups of interacting researchers) based on bibliometric data. The crucial question then is how much below the "tip of the iceberg" (recalling David Edge's metaphor) can we go when analysing bibliographic publication data for assessing research collaboration.? The best strategy is to use co-authorship analysis for obtaining a reasonable overview of the collaboration among groups.

The fundamental theoretical and empirical problem in this case was the identification of collaboration structures within the field of computational speech and language. If possible, the selection of researchers and research groups should be concentrated in the European area.

Operational Research Questions and Hypotheses

The research question can be formulated in this way: To what extent can we identify collaboration among groups within interdisciplinary fields of research, based on bibliometric data and co-authorship analysis? An inter-related question concerns the detailed identification of European research groups within the set of collaborating researchers identified by the longitudinal analysis.

Four key research topics have been pursued in order to determine a solution to the above questions. The first is the analysis of methodological strategies in order to overcome the problem of mapping the inter-disciplinary field of speech and language research. An "extended broad search" strategy based on a combination of several searching procedures has been utilised. Among these methods are "keyword search", "core journals" selection, "selected authors'" co-authored documents and "citing documents" to prolific authors.

Secondly, the importance of key contributors to the field of research (namely ASR) was tested against the empirical bibliometric data. The main issue here was that these outstanding contributors on the "collaboration map" of their field of research would have a major "structural impact". The evidence is focused on Rabiner's analysis in his seminal work of 1989. Their importance contributes to the identification process of more contributors to the field of research.

Thirdly, assuming that these systems are dynamic, this hypothesis focuses on identifying the dynamics of co-authorship networks, using a longitudinal analysis to illustrate research collaboration dynamics. Of particular concern here was the use of methods that allowed the dynamic mapping of these collaboration networks. Moreover, the bibliometric data should reveal the typical "speciation" process of clustering of research groups in their respective research specialties. Use of these methods should yield a very broad overview of the formation and evolution of speech and language research.

The final aspect was the identification of research groups based on the diverse collaboration maps. In particular, the feasibility of identifying European research groups within these "collaboration maps" was tested.

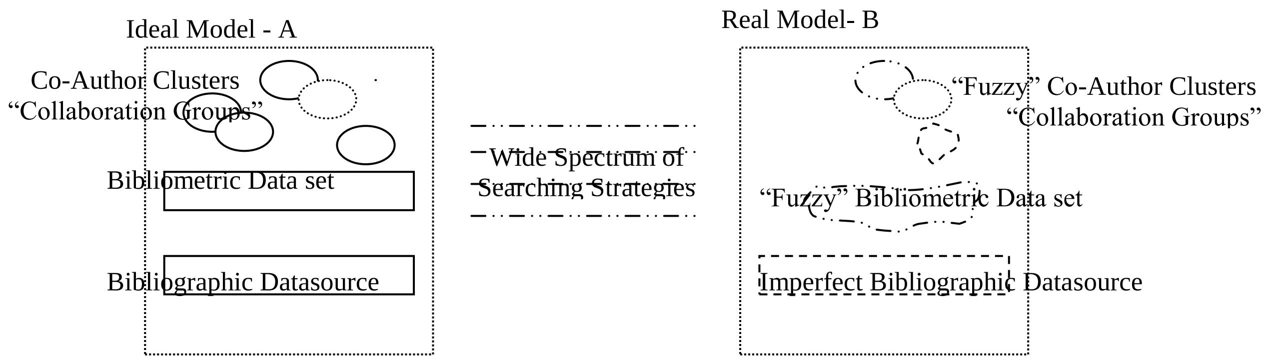
4.1.4 Mapping the Structure of Inter-disciplinary Fields of Research

This section focuses on the theoretical and empirical problem of how to delimit the boundaries of inter-disciplinary fields of research such as computational speech and language. Accepting from the outset that there is no optimal strategy for solving this problem, a second-best strategy was sought.

The delimitation problem, based on bibliometric methods, can be decomposed into two main stages. First, a bibliometric data set should be strictly defined as a base for the subsequent co-authorship analysis. This bibliometric set should ideally contain the whole set of publications in a given field. In the case of the current research it includes the whole set of publications in speech and language. At this stage the essential problem is the delimitation of the bibliometric data set, and particularly the criteria for such a delimitation. Second, there is the problem of identifying the structures of collaboration, as given by the bibliometric data set (and hopefully representing a significant proportion of the whole field of research). The problems here are related to the reliability and robustness of the co-authorship analysis, and mostly are dependent on the first stage.

It should be noted at this point that as part of a broad bibliometric approach, several other strategies for delimiting the boundaries of the research field should be pursued when feasible. Among those strategies, the most common are the use of specialised *Abstracting data sources*, the use of *Directories of researchers* currently active in the field, the use of *Publication lists* as given by research departments working in the field identified by "experts" and the use of *Rosters of researchers* involved in research projects or other bounded research activities. For delimitation of the European Speech and Language community based on research projects, see section 4.2. All these techniques would complement the bibliometric analysis based on more general bibliographic data sources, such as the ISI Science Citation Index abstracting databases. Being a relatively economical and practical source, these bibliographic databases are widely used for bibliometric analyses. Figure 4-1-1 depicts the essential characteristics of the bibliometric delimitation problem, allowing the representation of a wide spectrum of strategies for its solution.

Figure 4-1-1 - Delimitation of Bibliometric Co-authorship Analysis in Inter-disciplinary Research



In the diagram, the ideal Model A depicts the perfect scenario. In this case, the bibliographic data source (*e.g.* SCI) would cover the whole set of publications in speech and language research, as well as providing a powerful mechanism for retrieving inter-disciplinary research publications, based on some ideal classification system. This would allow, for instance, the selection of the whole set of publications in the sub-speciality of "Speech Synthesis" or "Speech Recognition". From this extensive and inter-disciplinary bibliographic data source, we would be able to construct an equally extensive, and well-defined bibliometric data set, that would be the basis for the co-authorship analysis. Under these circumstances, it would then be feasible to delimit the field of research reasonably clearly, and to identify reasonably delimited clusters of co-authorship. Those clusters, to a large degree, would represent the more dense groups of research collaboration, as given by formal publication.

Model B, on the other hand, depicts the actual scenario, that more commonly found in real explorations of bibliometric data. In this scenario, neither of the bibliographic data sources are complete, perfectly representing interdisciplinary publication, nor are the derived bibliometric data sets well bounded. The bibliographic data sources are imperfect, in the sense that they neither cover the whole set of journals for the field of research, nor do they provide a perfect "classification" mechanism allowing the selection of specialised publications. As a consequence, the resulting subsets of publications retrieved from the bibliographic data sources are commonly characterised by "fuzzy" boundaries. Naturally, the robustness of the latter co-authorship analysis and the identification of collaboration clusters is likely to be limited, in terms of the number of clusters as well as the boundaries of those same clusters.

An infinite number of intermediate search strategies is likely to be devised across this wide spectrum, in order to approximate the actual methodology of using co-authorship to characterise collaborating groups, to the optimality of Model A.

Within the field of information science, the problem of delimiting the boundaries of the bibliometric data set, as exposed above, can be modelled as a typical information search and retrieval problem. The objective is the maximisation of two variables: on the one hand, the *precision* or *accuracy* of the search and, on the other, the *coverage* or *breadth* of the search. Unfortunately, this represents a complex problem, as the maximisation of one variable (dimension) is usually at the expense of the other variable. In our research, this means that in most cases, the inclusion of a larger set of documents (journal articles) in the bibliometric data set - allowing a potentially more extensive representation of the field of Speech and Language - lowers the "precision" or accuracy of the bibliometric base (we end up having more false positives or publications that apparently have nothing to do with Speech and Language). An alternative strategy would be to select a more focused initial subset of publications, increasing the precision and accuracy results, but lowering the coverage of the publication data set used for subsequent co-authorship analysis. In empirical explorations, an example of the first strategy is the typical "Keyword Search" method, based on the selection of articles whose title, keyword and abstract fields contain a certain limited set of keywords. An example of the focused strategy is the traditional "Core Journals" method, based on the identification of publications drawn from a limited set of specialised journals in the field. Table 4-1-I shows the empirical results for speech and language for the period 1981 - 2000, based on the "Keyword search" method. As expected the coverage of a large number of publications was significant, but the precision of the results was affected. The number of false positives was about 30% of the whole data set - as assessed by a non-expert in the field.

However, it should be stressed that the "core journals search", while providing a more focused and *precise* result, has failed to identify significant contributions to the field. As an example, using Automatic Speech Recognition, Rabiner's seminal work, published in 1989 in the journal *Proceedings of the* IEEE, would have not been selected for the bibliometric data set, as *Proceedings of the* IEEE is outside the specialised "core journals" in the field. This type of problem is most likely to occur in highly dynamic, still maturing fields of research, where more general foundational journals might still represent the embryonic core of the research (this is certainly the case for ASR until 1993-1994). Among the several possible strategies to overcome the problem of p*recision* versus *coverage*, a broad search methodology, combining these various search techniques, has been suggested as the best approach. The literature review in the previous section (see, in particular, the last paragraphs of section 4.1.2) has briefly described some of these broad search methods, two of which are worth noting.

Persson and Beckmann (1995) suggested the use of a "core author" methodology. Publications "related" to a fundamental contributor to the field of research should be used as the main search method for defining the bibliometric data set. "Relation" is defined here mainly as "co-authored" and "co-cited".

Table 4-1-I - Publications in Speech and Language Research, ISI - SCI (1981 - 2000) - Keyword Search

Years	Speech Recognition	Speech Synthesis	Language Understanding	Language Processing
1981	16	14	3	11
1982	29	16	4	9
1983	46	22	6	21
1984	34	13	3	15
1985	41	7	1	25
1986	26	6	1	21
1987	36	6	0	18
1988	23	7	1	22
1989	30	7	2	21
1990	45	8	4	27
1991	171	26	38	241
1992	166	25	73	245
1993	162	37	88	323
1994	218	35	84	260
1995	272	40	142	393
1996	274	28	125	373
1997	285	34	116	388
1998	357	59	125	469
1999	393	36	152	553
2000	419	37	164	582
TOTAL	**3 043**	**463**	**1 132**	**4017**

Legend: Publications retrieved from Web of Science ISI - SCI bibliographic database

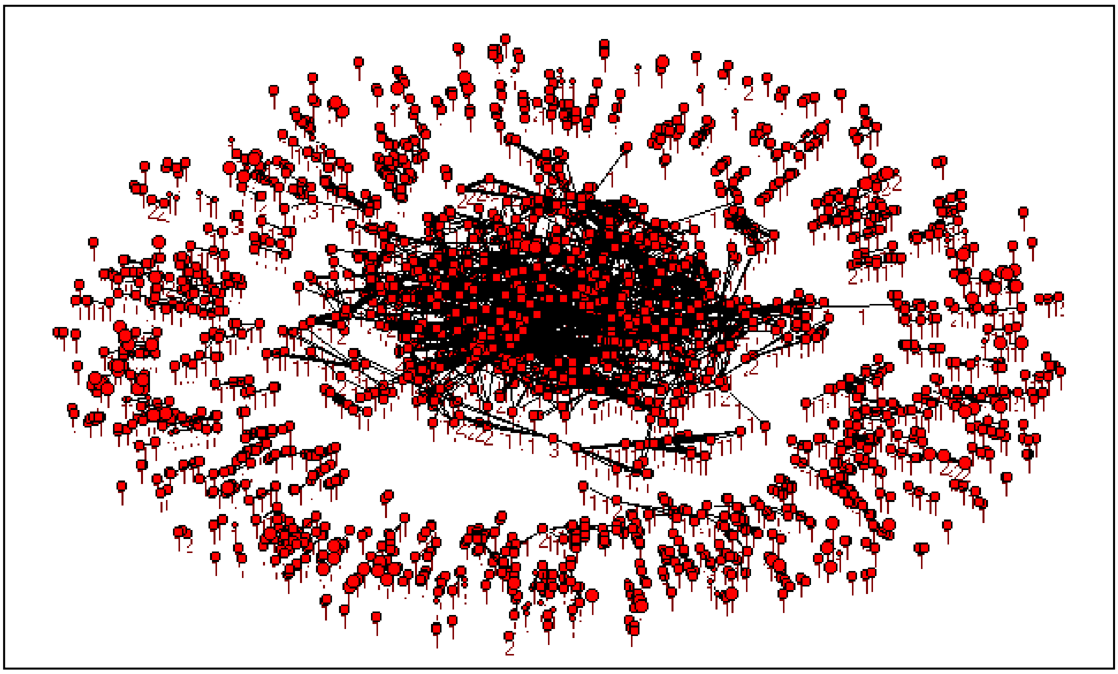
Using a more extensive strategy, Zulueta and Bordons (2001) suggested a search for delimiting the boundaries of a multidisciplinary field. The broad search strategy was based on Title Keywords, Institutional Addresses and Core Journals. In their case, the delimitation of the field of Cardiovascular Research, using the SCI classification categories was considered inadequate. With "Speech and Language" research, we have a similar situation. The SCI Computational Linguistics category and respective subcategories do not provide the solution for delimiting the boundaries of the base bibliometric data set.

In this present study a new methodology is suggested, based on a combination of the strengths of previous strategies. The Extended Broad Search method allowed a significant extension of the *Coverage* of the Speech bibliometric dataset (from an initial figure of 3,506 publications in both Speech Recognition and Synthesis based on the Keyword Search alone, to an eventual total of 5,925 combined with the Core Journal set). An additional set of 2,315 publications was found in the Core Authors search set. From this set 641 publications were repeated in the previous bibliometric set. The remaining subset was used to complete the Extended Broad Search set for Speech bibliometric analysis. For language research, a less extensive set of 362 publications was considered based on the Core Journal Set (covering the period 1995 - 2001). In structural terms, the extended coverage of the bibliometric data set allowed the identification of an additional set of collaboration groups (defined as cliques - see section 4.1.5 on the evolution of networks), estimated in at least 40% of the initial set in Speech and 15% in Language.

The assessment of the impact of the methodology in terms of the *precision* of the results would require a broad assessment by "experts" of the whole data set, which was outside the scope of the current investigation. Figure 4-1-2 graphically depicts the whole co-authorship network for ASR for the period 1999 - 2001. The collaboration groups identified for that period are discussed in section 4.1.7.

Figure 4-1-2 - Collaboration structures identified in ASR, using Extended Broad Search method

Publications (1999 - 2001) - 1,026 publications and 2,321 different authors
Resembling a Core - Periphery model of organisation -



Based upon the analysis of the bibliometric data set for this period, a matrix of co-authorship relashionships was built, relating each author with the whole set of other authors of publications gathered in the data set. Groups of authors co-authoring articles more frequently were aggregated. The quantitative analysis of the frequency of collaboration (based upon co-authorship of publications) and its graphical representation allows different groups of authors, contributing collaboratively more or less frequently in terms of publication, to be identified.

The graphical representation of the co-authorship map, based on a multi-dimensional scaling algorithm, that differentiates authors from their similarity/dissimilarity in terms of frequency of collaboration is represented in the figure above. The figure seems to show a "core-periphery" structure with a "core" of intensely connected (highly collaborating researchers) distinguishable from a "periphery" of researchers, who are less intensively connected. The next section discusses a similar result from assessment of the structural impact on the whole network of prolific "high impact" authors. The discussion is focused on ASR only.

4.1.5 Structural Impact of Revolutionary Contributions

"The Winners 'Structure' It All" Hypothesis

In this section we focus on analysis of the structural impact of significant scientific contributions. We analyse this "structural impact" in terms of the effects that seminal work, and outstanding researchers, have upon research collaboration in the field of research. The theoretical hypothesis underlying the discussion is the extent to which "revolutionary contributions" and the associated "prolific authors" are important for the overall structure of the research community. The empirical evidence is based on bibliometric data for the research speciality of "Automatic Speech Recognition". We analyse in particular a very significant contribution from Rabiner in 1989, that had a tremendous effect in advancing knowledge for solving the "speech recognition" research problem.

The discussion is organised in the following way. First, we present some historical evidence, corroborating the importance of Rabiner's article in this field. This historical evidence is set against some preliminary bibliometric indicators revealing the importance of his seminal work. Secondly, we discuss the "structural" impact, based on an analysis of three complementary indicators: 1) the longitudinal analysis of co-citation data and the resulting effects on research collaboration networks; 2) the extension (in time or number of collaborators) and associated structure of personal collaboration networks of the outstanding contributors; and 3) the enduring "citation" effect on other researchers in the field, as well as these structural characteristics.

We will first discuss the empirical and historical relevance of Rabiner's contribution. Within the area of "speech research", the speciality of "speech recognition" is focused on a specific but complex research problem: "How to convert an acoustic wave into a digital signal?" The solution to this problem had and will continue to have a major impact on several diverse applications (in research and technology), such as human-computer interfaces, telecommunication devices, remote automated controls, to name but a few. The solution to the problem is firmly rooted in the fields of digital signal processing, pattern recognition, electronic engineering and computational mathematics. The research efforts in these areas date back to the 1960s and 1970s with pioneering work at IBM laboratory research centres in the US on DSP, and in the late 1970s at Bell Laboratories also in the US.

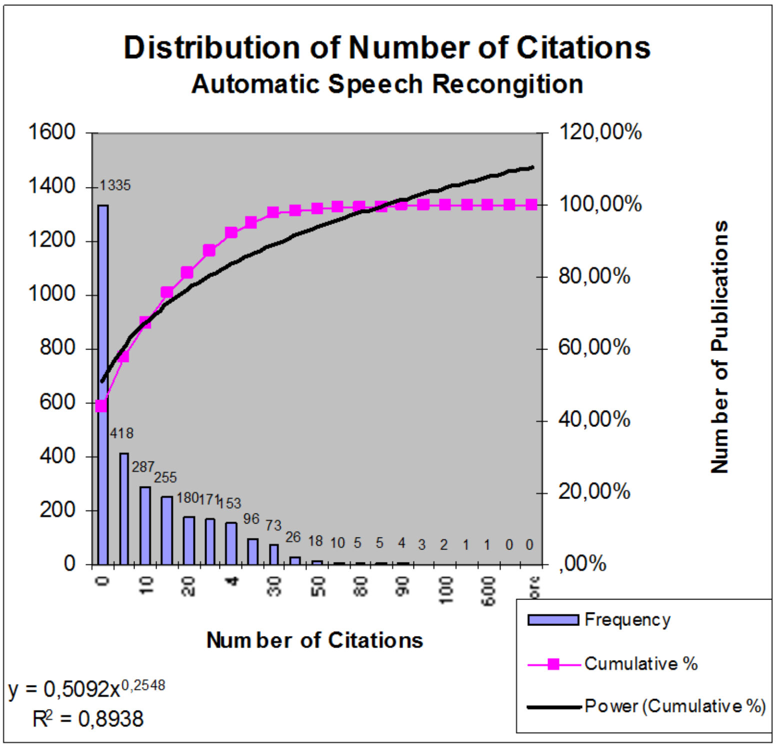
The significant "revolutionary" achievement of Rabiner, which took on public form with the publication of "A Tutorial on Hidden-Markov Models and Selected Applications in Speech Recognition", in 1989, in the *Proceedings of the IEEE*, was to bring a mathematical method/technique and integrate it into the problem of signal recognition. In itself, this seminal work represented the inter-disciplinarity and trans-disciplinarity[28] of this field of research. Since this outstanding contribution, the field of speech recognition has flourished and evolved at an astonishing pace. In the late 1990s, the state-of-the art in the "speech and language" research area reflected the maturity of this specific speciality, recognising that, even if not completely problem free, the results from applying this research to application-oriented technologies, available on the market were quite remarkable (Cole *et al.*, 1997).

The effective relevance of Rabiner's work should to a certain degree be identifiable in the bibliometric data. First, we briefly analyse the "impact" of Rabiner in terms of his "prestige" within the whole network of researchers. In a second stage, we analyse the structural impact of his contribution.

A crude, but widely used, measure of *prestige* in bibliometric terms is achieved from the total *number of citations* by *peers* to a certain author's work. Moreover, the extension over time of those citations is also an important indicator of long-lasting "impact". The bibliometric data identifies Rabiner's 1989 article as the most cited document and Rabiner as the most cited author within the ASR community, amongst a small number of other highly cited authors. These results are significant in a context of a very unequal distribution of the citations over the whole set of publications (corroborating an inverse power law distribution, widely identified in the literature since Lotka's work in the 1920s and Price's later 1960's work. Figure 4-1-3 shows the distribution of citations per publication.

[28] We adopt here a conceptual distinction between multi-disciplinarity, inter-disciplinarity and trans-disciplinarity. An activity is *multi-disciplinary,* when it assembles, in an additive fashion, knowledge from more than one discipline. On the other hand, an *inter-disciplinary* activity, is one that produces knowledge that integrates more than one discipline. Integration is the defining element. *Trans-disciplinary* activities result from activities conducted by individual researchers or research groups that cross several disciplines in the course of the research work. Part of this conceptualisation is discussed in Gilbert (1998).

Figure 4-1-3 - Distribution of Citations in "Automatic Speech Recognition" - 1981 - 2000



In addition, Table 4-1-II identifies the 10% most cited articles in the whole set. It should be stressed that some of these articles were not published in "core journals"; however, they were published in foundational journals in the field of ASR, reflecting the fact that digital signal-processing represented one of its main disciplinary origins.

Another significant result revealed by the bibliometric data, is that at least two of the most cited authors (Rabiner and Juang) were from ATT Bell Labs, and a third (Bahl) was from the IBM Corporation (Thomas J. Watson Research Centre, US), these being two of the most important research centres in ASR, over the years.

Table 4-1-II - Relevance of the "most significant contributions"

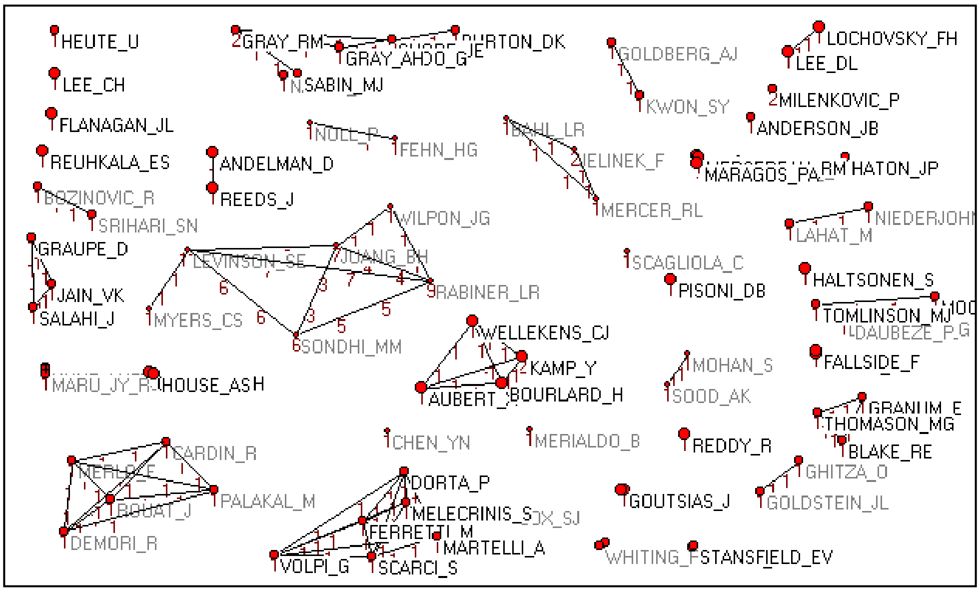
Year of Publication	Journal of Publication	Times CITED
1989	PROCEEDINGS OF THE IEEE	521
1983	IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE	158
1987	JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA	90
1984	JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA	88
1992	IEEE TRANSACTIONS ON NEURAL NETWORKS	85
1986	IEEE TRANSACTIONS ON ACOUSTICS SPEECH AND SIGNAL PROCESSING	78
1992	IEEE TRANSACTIONS ON SIGNAL PROCESSING	58
1987	IEEE TRANSACTIONS ON ACOUSTICS SPEECH AND SIGNAL PROCESSING	53
1992	JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA	51

The above bibliometric results provided a general indication of the "impact" of core authors on the whole research field and network of researchers. Nevertheless, the relevant test for our purposes is the "structural" impact of seminal work by these "core authors". The empirical analysis focuses on Rabiner's seminal article in 1989. The empirical evidence is based on: 1) the structural impact in terms of the number and connectivity of the network members, as given by the "knowledge-map" of the field; 2) the extensiveness of the direct personal network of Rabiner - the researchers with whom he collaborated and the direct collaborators of his co-authors; and 3) the enduring "citation impact" of his work. We analyse the results in the following paragraphs.

Structural impact on the "Knowledge-map" of "Automatic Speech Recognition"

Figure 4-1-4 and Figure 4-1-5 depict the co-author network based on bibliographic-coupled documents[29] of ASR, respectively before and after the publication of Rabiner's 1989 article.

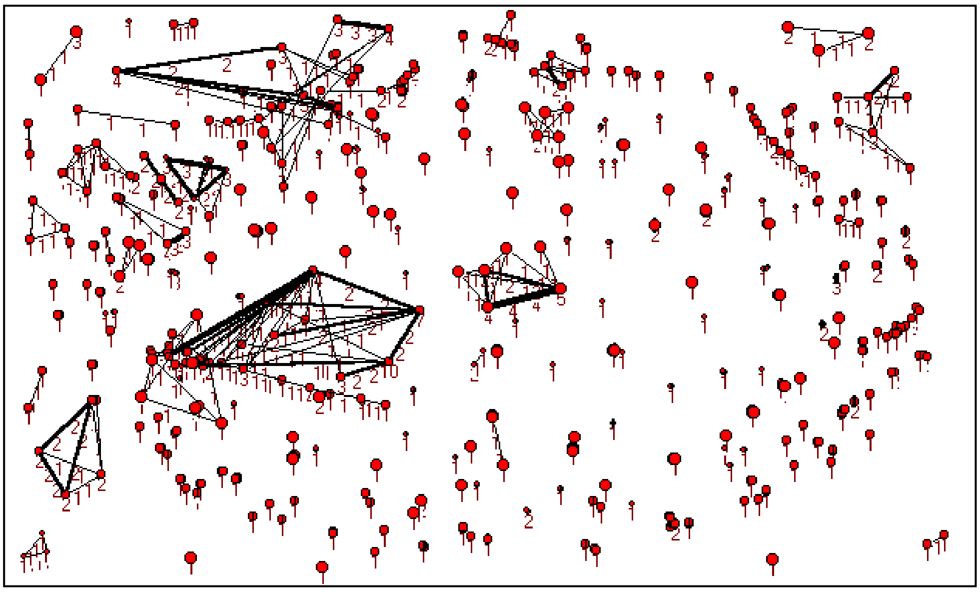
Figure 4-1-4 - Co-author Network before seminal work (1981 - 1988)



The co-author maps were constructed based on the bibliographic coupled data sets. In Figure 4-1-4 only one clique of collaborating authors is shown (the definition of a clique is based on the restriction that authors should have between them and within that clique more than one co-authored paper. The "collaboration group" was based at ATT Bell Laboratories (Bell Tel. Laboratories, at that time) and comprised Rabiner, Levinson, Sondhi and Juang (all of whom are highly cited authors in this field).

[29] It should be noted that these "Knowledge-maps" were not constructed under the assumption of co-citation analysis, but based on bibliographic-coupled documents. Two documents are considered to be bibliographically coupled if they share one or more bibliographic references. This is a "weaker" assumption than co-citation which requires two documents to be cited simultaneously in other documents.

Figure 4-1-5 - Co-author Network after seminal work (1989 - 2000)



An immediate first structural impact of seminal work is likely to be the strengthening effect on the collaboration group of the author (e.g. through recruitment of new members, allocation of funding for new projects, etc.). In fact, the structural analysis allows us to verify the extension of the ATT Bell Labs' research group, in terms of number of researchers, the connectivity links among them, and the strength of association as given by published output. The co-authorship analysis for the second period (1989-2000) identified 11 intensely collaborating groups. The vast majority of these groups are now recognised as important "nodes" in the international community of Automatic Speech Recognition. *Annex II* presents the full list of "collaboration groups". Some of the groups identified are the ATT Bell Labs, SRI International - Speech Technology Research Lab, Duke University - Department of Electrical and Computer Engineering in the USA, University of Toronto - Department of Electrical and Computer Engineering in Canada and the City University of Hong-Kong - Department of Computer Science and Department of Electronic Engineering, in China).

It should be noted that this structural effect on the "Knowledge-map" is not restricted to Rabiner. A close inspection of the "bibliographic coupled co-authored networks" of the other 7 most cited authors (Bahl, Vantasell, Dubno, Pal, Furui, Juang, and Meddis), shows up similar structural effects (more researchers, greater level of connectivity and strengthening of association).

The Extensiveness of the Personal Collaboration Network of Prolific Contributors

A second indicator for assessing the structural impact of Rabiner's work is provided by the characteristics of his immediate "neighbourhood" of collaborators. The effect of Rabiner's prolific work is illustrated by the extensiveness of his collaborating network, the clustering effect on other parts of the network and the importance of the "nodes" linked to Rabiner.

Figures 4-1-6 and Figure 4-1-7 map the first-order collaboration network of Rabiner (i.e authors identified from his individual and co-authored articles), and the second-order collaboration group (i.e. the immediate collaborators of Rabiner's collaborators).

Figure 4-1-6 - Ego-Network of Rabiner (direct collaborators)

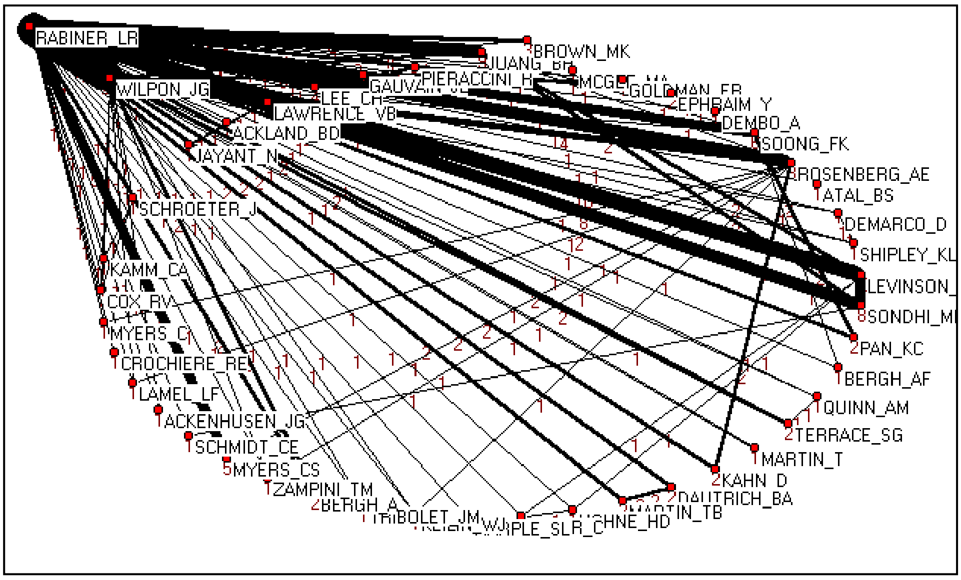
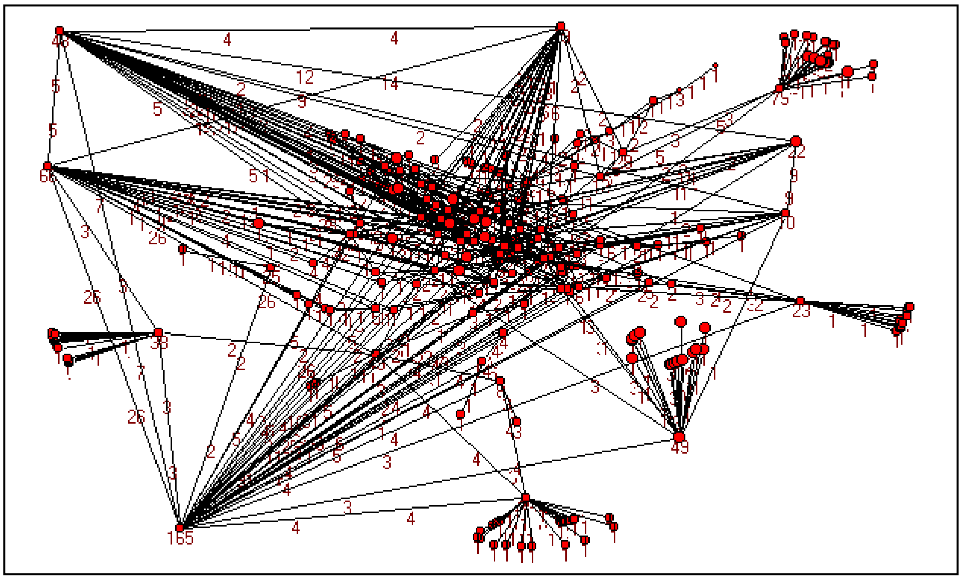


Figure 4-1-7 - Ego-Network of Rabiner (second layer of co-authorship)



The extensive number of immediate collaborations with prolific authors is an approximate indicator of the "activity" and "prestige" of such author. Even more important is the "structuring effect" of this prolific collaborating network, i.e. the effect on the strength of the strong and weak ties (Granovetter, 1973). Rabiner's work was important not only because he produced numerous papers and was frequently cited (the previous empirical evidence testifies to this), but also because it helped to establish new "important nodes" of research in ASR all around the world. An analysis of the neighbourhood of collaborators allows identification of the researchers who now lead research centres in ASR in such geographically dispersed countries as Portugal (INESC), France (LIMSI), China (City University, Hong Kong) and Canada (University of Toronto).

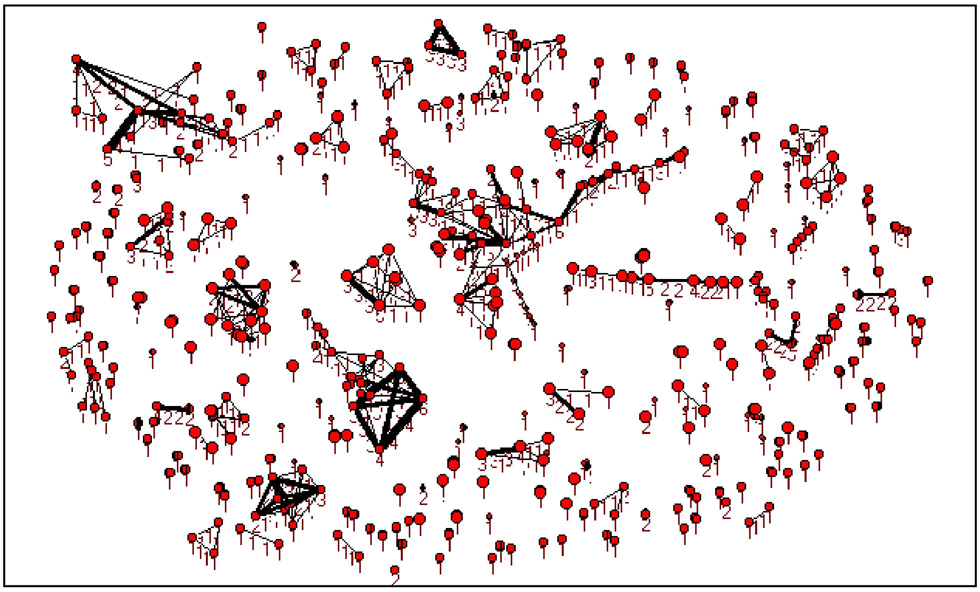
The work of prolific authors supports the structuring of these more intensive collaboration groups, as is evidenced by Figure 4.1.7. This reveals the "importance" in terms of connectivity and intensity of linkages not only with Rabiner (bottom left), but also with some of his immediate collaborators (e.g. Wilson at the top left and Juang at the bottom right with 49 links). These collaborators generate their own structural linkages and strengthen their own immediate neighbourhood of collaborators. Nevertheless, it should be stressed that each "nodal researcher" possesses a more intense and tightly bounded vicinity - this is better illustrated by Figure 4-1-6 where more intense links are indicated by thicker lines.

A third indicator analysed here, assessing the structural effect of "core authors", is represented by the structural "citation effect".

Enduring "Citation Network" of Outstanding Contributions

The enduring citation effect of a seminal work is also corroborated by the extension over time of the citations to that work. This effect also has a structural counterpart, i.e. the diversity and extension of the "collaborating groups" that, over time, cite the seminal articles and other frequently cited research work. Despite the fact that Rabiner's seminal article was published in 1989, in the period 1994 to 2000, a total of 931 different authors cited this work. Moreover, those authors were identified as being "organised" in more or less intense "collaboration groups" detected by co-authorship analysis. Figure 4-1-8 depicts the co-authorship network based on the citing documents.

Figure 4-1-8 - Co-authorship network of documents citing Rabiner's seminal article



The co-authorship analysis of this "citing network" allowed the identification of 25 different cliques (dense and intense collaborators). In Annex III a full list of these "collaboration groups" is provided. The research groups identified in Europe include Siemens Brussels, Advanced Processing Control Group, Belgium; University Politecnic of Catalunya, Dept Signal Theory & Communication, in Spain; University of Bremen, Centre for Cognitive Sciences, in Germany; and Université Paris 07, INSERM, U155, Equipe Bioinformatique Molecular, in France.

However, the structural "citing impact" is geographically dispersed to include the Australian National University Department of Systems Engineering; the Hebrew University of Jerusalem Institute of Computer Science in Israel; the University of Windsor Department of Electrical & Computer Engineering in Canada; and the Kyushu Institute of Technology Department of Artificial Intelligence in Japan.

The above empirical evidence confirms the hypothesis that "core authors" have a significant structural effect on their fields of research. This bibliometric assessment is valid whether we measure that structural effect in terms of size of collaborating networks, diversity of those collaborating networks, the citing structural effect over a reasonable period of time, or more generally by the formation and evolution of the field of research, as revealed by the changes in the "knowledge map" of the field. In a more provocative assessment, we seem to have some validation for the hypothesis that "The winners 'structure' it all". The next section looks in more detail at the importance of co-authorship analysis for mapping the formation and evolution of fields of research in structural terms.

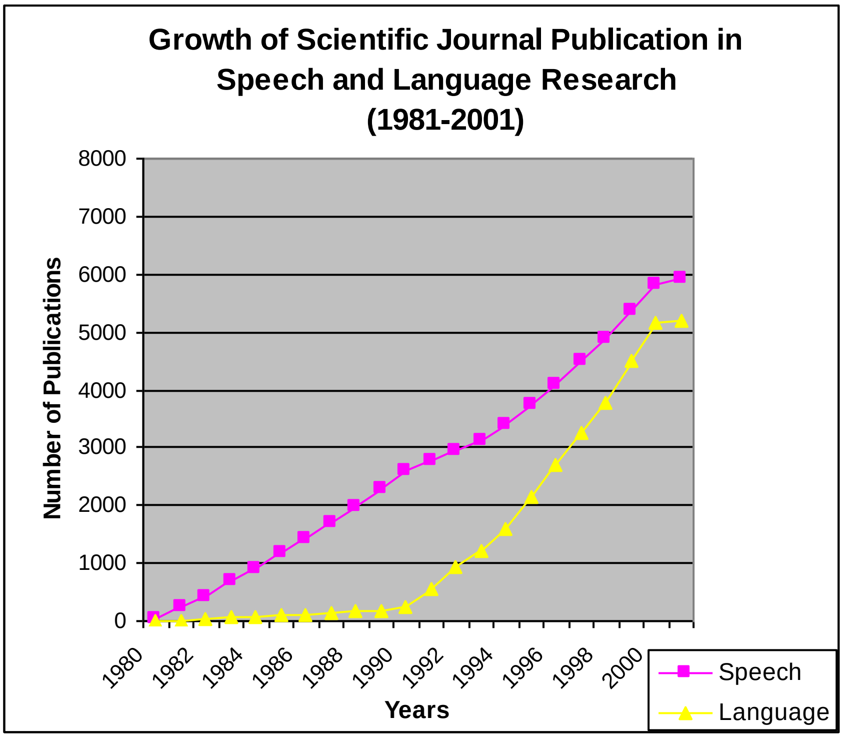
4.1.6 The Formation and Evolution of Research Collaboration Structures

This section discusses the hypothesis that the indepth analysis of co-authorship networks identified through longitudinal analysis of publication, reveals the formation and evolution of collaboration structures within a field of research. Empirical evidence on co-authorship networks for two different "communities" was analysed (on the one hand, speech research, and on the other hand, language research). Notwithstanding the fact that the fields of research are different, and also that there are specialities and sub-specialities within the fields of research, important regularities in structural patterns of evolution are detected and can be analysed. Some methodological procedures are also suggested for mapping such structural evolution.

Theoretical knowledge of the research area of "Speech and Language" suggests that these two fields are at different stages of evolution. This background knowledge was obtained by a review of the literature in those fields (particularly the most recent assessment of the state-of-the-art in Speech and Language (Cole *et al.*, 1997), and by consultation with two "experts".

The bibliometric data appear to validate this hypothesis. A first approximation to the evolution of these fields is given in Figure 4-1-9, which depicts the cumulative growth in the number of publications.

Figure 4-1-9 - The Exponential Growth of the Scientific Literature - Speech and Language Research



The growth stage for Language research apparently started in 1992 and, globally, the Language research field has followed a typical "exponential growth" model, whereas Speech research adopts a smoother evolutionary pattern, beginning much earlier in the mid-1980s and showing some acceleration during the 1990s. An analysis of the collaboration structures of both areas validates this assumption that these fields are indeed different communities. The identified clusters are different for the two areas, but with some research groups "bridging" the two fields. This led to a separate structural analysis of the evolution of each research field. The following discussion is based on empirical evidence from Language research. We briefly introduce some theoretical assumptions underlying the analysis and then discuss the empirical results.

Scientific communities evolve over time as does the knowledge base generated by their research activity. The "quasi-exponential" growth of knowledge (illustrated above in Figure 4.1.9) is paralleled by the changing nature of the structure of the scientific collaboration network (see Figures 4-1-10, 4-1-11 and 4-1-12 depicting the collaboration maps) in the course of the evolutionary process.

Scientific productivity, as measured by the "proxy" of bibliometric indicators, is related to the (structure) pattern and topology of the network. Indeed, the speech recognition data reveal that different components of the network show different levels of productivity and this simple fact dramatically influences the subsequent evolution of the structure of the network (we analysed this topic in the previous section). Moreover, the diversity of specialised research programmes within the network is an interesting property, with growing patterns of "speciation", and subgroups focusing their efforts in subspecialties of the "Language research" problem. This testifies to the "division of labour" characteristic of the scientific process, as well as to an evolutionary adaptation of social subgroups (subspecies) - or closely connected subgroups of scientists - to the environment, selecting a specific "ecological niche" in which to conduct their research.

The fundamental assumption is that the structure of collaboration networks, as given by co-authorship analyses, reveals the evolution of the field. The empirical findings validate this hypothesis. To analyse the dynamics of the collaboration structures, a methodology combining a continuum approach with a cross-sectional approach was devised. Only a combination of the two methods allows a full understanding of the evolution of Language research. While the *Continuum approach* allows a thorough analysis of the marginal dynamics of the structure of the field from one period to the next the *Cross-sectional* approach allows a structural representation of the field during a certain time-frame. It constitutes a snapshot or static representation of the field at that time. This is useful for comparing the two research fields of Speech and Language. Table 4-1-III summarises the indicators for each approach in the empirical analysis of Language research, over the period 1981 - 2001.

Table 4-1-III - Summary of the indicators for analysis of the Structural Evolution of Language Research

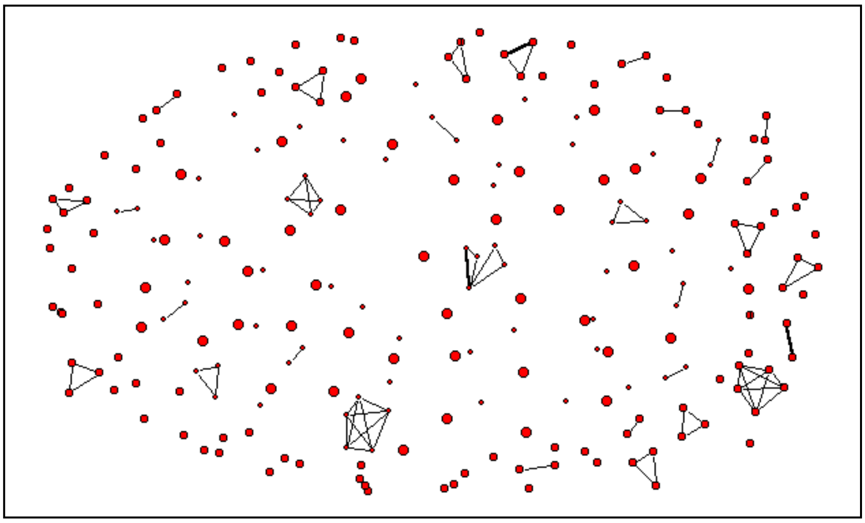
Language Research Structural Evolution - Bibliometric data set (1981 - 2001)					
Type of Analysis	Years	Number of Publications	Number of Researchers	Groups Identified * 1 Collab	More than 1
	1981 - 1989	187	308	39	0
Continnum	1981 - 1990	250	432	58	1
	1981 - 1991	556	1103	156	3
	1981 - 1992	930	1898	277	14
	1993 - 1994	647	1693	277	8
	1995 - 1996	1191	2573	420	35
Cross-Section	1997	571	1545	-	7
	1998	605	1686	-	12
	1999	780	1955	-	23
	2000 - 2001	787	1900	-	19

The collaboration groups are identified by co-authorship analysis. A "collaboration group" corresponds to a clique in technical terms (a group formed by a minimum of three researchers who have co-authored a paper, groups being distinguishable by the intensity of those collaborations, represented by the number of co-authored publications). For the structural analysis the relevant groups are those with the most collaborations, indicating the strength of "collaboration". Nevertheless, over time, the growth of single paper cliques, in terms of the number of groups, as well as the number of members within each clique, is a good indicator of the growth of collaboration within the field.

The *Continuum approach* is thus particularly suited to analysis of the evolving structure of the collaboration network. Structural evolution is analysed at three different levels: the patterns of connectivity for individual researchers; the connectivity of collaborating research groups (that might or might not represent institutional collaboration groups); and finally the patterns of connectivity of the whole network.

The empirical evidence for language research, over the period 1981 - 1992, was analysed using a continuum approach. This period was chosen on the assumption that it represented an "emergent" or "incipient stage" in the field. For later periods the cross-sectional approach was used. Figures 4-1-10, 4-1-11 and 4-1-12 depict the structural evolution of Language research.

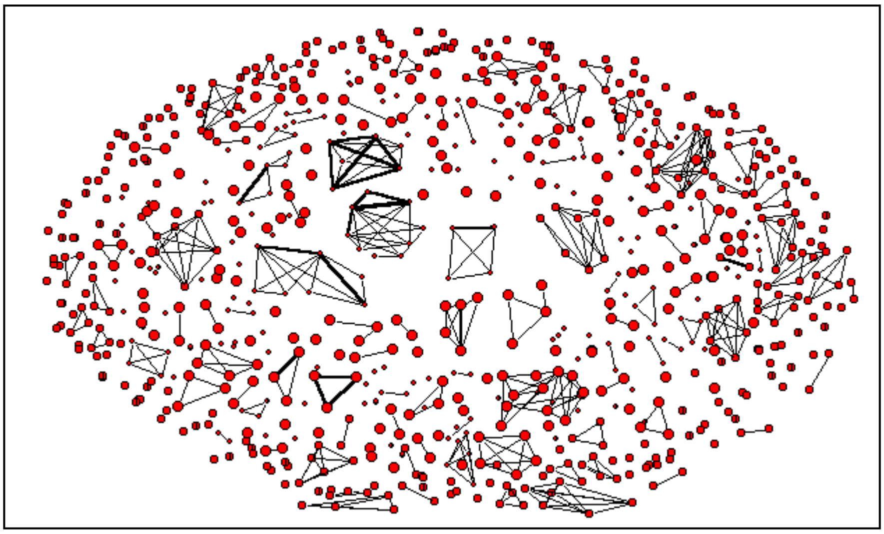
Figure 4-1-10 - Co-authorship Network - Language Research (1981 - 1989) - Incipient Stage



Despite covering a reasonable period of time (1981 - 1989), at this early stage of the formation of the Language research community, the number of publications, researchers and groups is limited. No single "collaboration group" was identified as having more than one collaboration. The whole network was sparse, disconnected and made up of a multitude of isolated groups and just a few co-authoring weak groups. This would therefore seem to correspond to the "incipient stage" of the field. If the period is extended to 1990 then a more intense collaboration group can be identified which was established by Renier, Eling, Slis and Debot (University of Nijmegen, Department of Experimental Psychology, The Netherlands).

As the number of co-authored publications increased, the number of co-authoring groups also grew. Simultaneously, the "intensity" of collaboration within existing collaboration groups was strengthening and the number of researchers in each group was growing. This systemic process is apparent if Figures 4-1-11 and (covering publications up to 1991) is compared with Figure 4-1-10.

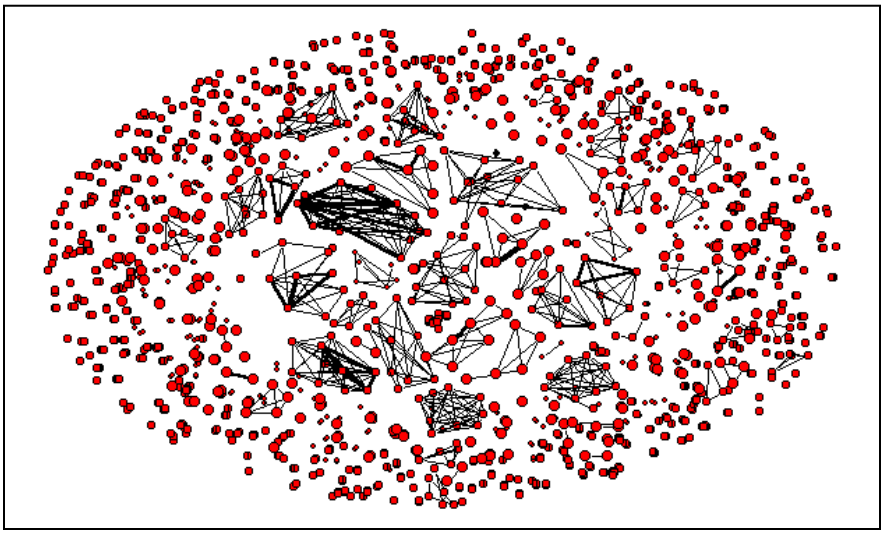
Figure 4-1-11 - Co-authorship Network - Language Research (1981 - 1991) - Incipient Advanced Stage



At this stage (here designated the "Incipient advanced stage"), a small number of more intensively bounded collaborating groups can be identified. One group was formed by Crookes, Morrow, and Mcparland, from Queens University of Belfast, Department of Computer Science, Ireland and another group consisted of Numazaki, Tamura, and Tanaka, from the Yokohama National University and the Tokyo Institute of Technology in Japan.

A significant transformation occurs when the network enters the "Growth stage". The natural increase in the number of members of the community is by this time significant. The same is true of co-authored activities and groups. The structural difference is given by the following characteristics: the number of intensively collaborating groups exponentially increases; these groups are enlarged, in terms of the number of collaborating researchers; and, finally, the intensity of collaboration within already existing groups is strengthened in a significant way. Figure 4-1-12 depicts the structure of the Language research collaboration network, at the beginning of the growth stage.

Figure 4-1-12 - Co-authorship network - Language Research (1981 - 1992) - Growth Stage – Beginning



Fourteen intense collaboration groups were identified by co-authorship analysis at this stage. A complete list is provided in Annex IV, with all the other Collaboration Groups identified by Co-Authorship analysis using the Continuum approach (some of these "Collaboration Groups" are discussed in the next section, section 4.1.7).

Further exploration of the evolutionary growth model is not attempted here, but the main structural dynamics have already been pointed out. A comparison of these structural dynamics with the results of the cross-sectional approach provided the background for a discussion of other topics related to the formation and evolution of collaboration networks.

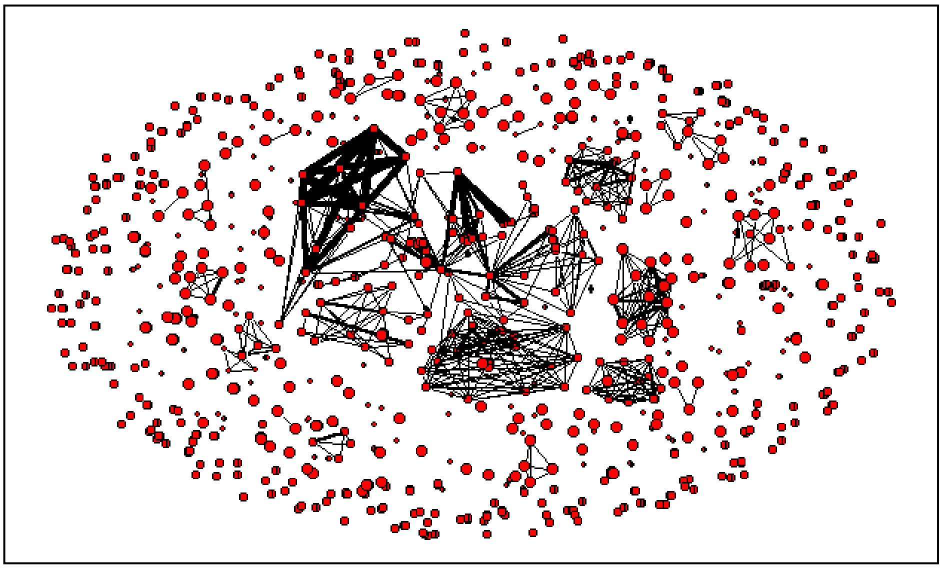
Analysing a snapshot of the collaboration structure introduces a different perspective from the one adopted so far. Despite this methodological difference, some results are worth commenting upon. First, the empirical result of the cross-sectional analysis over several consecutive periods provides additional evidence of intensively collaborating groups of researchers (see Annex VII for an extensive list of such collaboration groups).

Secondly, it seems that the identified structural collaboration groups are also different in terms of their "specialisation" or the nature of their research. The extensive analysis of the identified collaboration groups, the journals in which they publish and the titles of their publications, allows different "zones" of the network with different conceptual interests to be mapped. For example, a special area of application-oriented research on Medicine was identified with several research groups from radiology and pediatry; a different group was identified in computational aspects of linguistics; and still other groups in psycholinguistics and neuroscience. These results might reveal that the structural analysis is based on too broad a set of publications (which is methodologically a problem but at the same time provides a robust process for distinguishing research specialities within the same research field). On the other hand, these empirical findings might also reveal the evolutionary "division of labour process", with the fragmentation of the initial network into "*speciated*" collaboration groups, focusing on particular, more specific research problems.

Thirdly, the cross-sectional analysis also reveals that the structure of the whole network is different at more advanced stages of development of the field of research. Some of the important differences are as follows. First, the different collaboration groups interlink with one another. In fact, this is more likely to happen among very intensively collaborating groups. This is empirical evidence of inter-group collaboration. Secondly, the individual links of some researchers and some research groups affect the connectivity of the whole network. It is noticeable that some "bridging" authors inter-link different collaboration groups, allowing an increase in the density of the whole network. Thirdly, the structure of the network nevertheless reveals a high degree of disconnectedness. Two important components are clearly identified: a "core" which includes the more intensive inter-collaborating groups; and a "periphery" of more dispersed, less interconnected groups and individual researchers. A particular reason for the identification of a "core" with more "productive" researchers and research groups is that more "active" researchers are more likely to have more connections with other researchers. However, this does not necessarily mean more inter-group connections. The empirical evidence shows that most of these connections are inside a well bounded, tightly connected group of collaborators. This applies to all but a few researchers, who are the "bridging" researchers or "research troubadours" of Language research.

Figure 4-1-13 depicts the results of the cross-sectional analysis of Language research in the period 2000 - 2001, allowing the identification of some of the characteristics discussed above.

Figure 4-1-13 - Co-authorship Network - Language Research (2000 - 2001) - Cross-sectional Growth Stage



The cross-sectional approach also allows the structural comparison of two research fields. In this case, a visual inspection of Figure 4.1.2 and Figure 4.1.13 depicting respectively the collaboration structures of Speech research and Language research, over approximately the same period of time, reveals quite different group structures. While both maps validate the "core" and "periphery" structure model, the Speech network reveals a greater level of connectedness within the "core" component (it shows more researchers, more research groups and more tightly bounded research structures). This was perhaps expected given a background knowledge of the differences in the evolutionary stage of each research area. This analysis needs to be supported by a more extensive evolutionary analysis of other cross-sectional periods, which this investigation has attempted to provide.

Some exploratory ideas are briefly outlined, before proceeding to the results in the next section. Regarding the evolution of the "core", the above results appear to confirm the findings of previous research (e.g. Zulueta and Bórdons, 2001) of a mixture of "fragmentation" and "hybridisation" among intensely collaborating groups. The "fragmentation" involves more "speciated" collaboration groups that gradually form their own research speciality and eventually a new field of research. The "hybridisation" comes about from the formation of even more inter-disciplinary collaboration groups. Nothing can be said about the structural evolution of the periphery *per se*, but a theoretical exploratory hypothesis would be that some individual researchers end up joining the "core" group, while the majority of them leave the network. Both these topics are subjects for further investigation.

In the next section, the identification of "collaboration groups" using the method of co-authorship analysis is described.

4.1.7 Identification of Research Collaboration Groups based on Bibliometric Data European Speech and Language Research

The empirical findings of previous sections leads to confirmation of the assumption that co-authorship analysis of bibliometric data does indeed allow the identification of research "collaboration groups". However, a refinement of this assumption is necessary. First, we need to agree on the exact meaning of a "collaboration group". Secondly, only the combination of a number of different approaches yields reliable results. And, thirdly, the identification process requires ex-post "expert" assessment or triangulation with other information sources on collaboration (such as implementation of a survey).

The concept of a "collaboration group" is based on the social network analysis definition of a *"clique"*. We have previously clarified this concept. Members of a group have links with all other group members of varying degrees of intensity (given by the number of links with the same member). The extrapolation of this technical term to other "forms" is necessarily subject to error. A first necessary refinement is that a "collaboration group" is not necessarily an "institutional research group". It might be formed by individual researchers working within the same institution but not in any institutionalised research group. Secondly, the bibliographic databases do not always clearly indicate the institutional and research group affiliation of each author of a paper (this has been extensively covered in Katz and Martin, 1997). For example, a group of three authors in a bibliographic record might have four institutional addresses and it might be impossible to cross-relate author affiliation with other records. Thirdly, inter-individual collaboration does not necessarily entail inter-group collaboration or inter-institutional collaboration or inter-national collaboration. The following example clarifies the problems inherent in the concept of "collaboration group".

"Collaboration Group" Identified in both Speech and Language data set (1981 - 2000)

4: OVIATT_S LEVOW_GA MACEACHERN_M
Oregon Grad Inst Sci & Technol, Dept Comp Sci, Ctr Human Comp;Commun, POB 91000, Portland, OR 97291 USA.;
Oregon Grad Inst Sci & Technol, Dept Comp Sci, Ctr Human Comp Commun, Portland, OR 97291 USA.;
Univ Pittsburgh, Dept Linguist, Pittsburgh, PA 15260 USA.;
MIT, Artificial Intelligence Lab, Cambridge, MA 02139 USA.

In this case, we have the latter two kinds of problem: we cannot reliably deduce the affiliation of each individual researcher. Potentially, the second author (Levow) has two institutional addresses - the Oregon Graduate Institute of Science and Technology and the University of Pittsburg. Or, it might also be that the third member of the group belongs to the University of Pittsburg and MIT. Only complementary information might resolve these technical problems. A different and more substantial problem is whether this specific case of inter-individual collaboration corresponds to inter-institutional collaboration. Nothing from the analysis of the numbers alone can confirm whether in fact all three institutions and research groups are effectively collaborating. However, long-term analysis of collaboration structures and the repetitive pattern of inter-individual collaboration of other researchers from the same groups may strengthen the results that emerge.

A final remark on the technical problems involved in identifying group members within "collaboration groups" concerns the inner dynamics of the system and the mobility of researchers. As most bibliometric analyses are conducted over long time periods, the institutional affiliation of group members often changes and the address in a publication is not necessarily consistent with a researcher's current address.

Taking into account the above clarifications of the concept of "collaboration group" and the limitations of extrapolations for the process of "research collaboration", the next requirement for reliable identification of research groups by co-author analysis is the combination of several complementary strategies. This situation is compounded when identifying groups in inter-disciplinary fields of research. Among the techniques available, it was necessary to use a combination of: 1) an Extended Broad Search method or similar approach for the delimitation of the bibliometric data set; 2) the separate analysis of bibliographically-coupled documents with important works in the fields of research; and 3) the long-term analysis of the evolution of the co-authorship networks.

The combination of these methods allowed identification of the groups within the European Speech and Language area. Some of those "collaboration groups" are listed in Table 4-1-IV, indicating some highly visible international groups.

Table 4-1-IV - Selected Research Groups Identified by Co-authorship Analysis

European Speech Research
Univ Amsterdam, Inst Phonet Sci, IFOTT, NL-1016 CG Amsterdam, Netherlands
Departamento de Sistemas Informáticos y Computación, Universidad Politécnica, Valencia, Spain
Univ Granada, Fac Ciencias, Dept Elect & Tecnol Comp, E-18071;Granada, Spain
Univ Nijmegen, Dept Language & Speech, A2RT,POB 9103, NL-6500;HD Nijmegen, Netherlands
Lab Linguist Formelle, Case 7003,2 Pl Jussieu, F-75251 Paris;France
Univ Paris 07, CNRS, F-75251 Paris, France
Univ Cambridge, Dept Engn, Cambridge CB2 1PZ, England
Philips Res Labs, POB 1980, D-52021 Aachen, Germany
Siemens Brussels, Adv Proc Control Grp, Huizingen, Belgium
CNRS, LIMSI, Spoken Language Proc Grp, BP 133, F-91403 Orsay,;France
Univ Bonn, Comp Sci Dept 3, D-5300 Bonn, Germany
Georgian Techn Univ, Dept Digital Commun Theory, GE-380075 Tbilisi, Rep of Georgia
Univ Surrey, Ctr Commun Syst Res, Guildford GU2 7XH, Surrey, England
European Language Research
Max Planck Inst Cognit Neurosci, D-04103 Leipzig, Germany
EHESS, CNRS, Lab Sci Cognit & Psycholinguist, F-75651 Paris 13, France
Univ Karlsruhe, AIFB, Kaiserstr 12, D-76128 Karlsruhe, Germany
Univ Erlangen Nurnberg, Chair Pattern Recognit, Martensstr 3,;D-91058 Erlangen, Germany
Bavarian Res Ctr Knowledge Based Syst FORWISS, D-91058 Erlangen, Germany
Univ Freiburg, Computat Linguist Lab, D-79085 Freiburg, Germany
MRC, Cognit & Brain Sci Unit, Cambridge CB2 2EF, England
Univ Nijmegen,Dept Exptl Psychol,6500 HE Nijmegen,Netherlands
Univ Maastricht, MATRIKS, Dept Comp Sci, Maastricht, Netherlands
DFKI, D-66123 Saarbrucken, Germany
Leiden Univ, Dept Comp Sci, POB 9512, NL-2300 RA Leiden,;Netherlands
Highly visible International Research Groups
AT&T Bell Labs, Res, Florham Park, NJ USA.;AT&T Bell Labs, Res, Florham Park, NJ USA
Stanford Univ, Dept Math, Stanford, CA 94305 USA
SRI Int, Speech Technol & Res Lab, 333 Ravenswood Ave, Menlo;Park, CA 94025 USA
Carnegie Mellon Univ, Dept Comp Sci, Pittsburgh, PA 15213 USA
Technion Israel Inst Technol, Dept Elect Engn, IL-32000 Haifa,;Israel
City Univ Hong Kong, 83 Tat Chee Ave, Kowloon, Hong Kong
Duke Univ, Dept Elect & Comp Engn, Box 90291, Durham, NC 27708;USA
Univ Toronto, Dept Elect & Comp Engn, Toronto, ON M5S 3G4, Canada
Tsing Hua Univ, Dept Comp Sci, Beijing, Peoples R China

The above list is not meant to be exhaustive. Annexes VI and VII provide a listing of all the collaboration groups detected in Speech research and in Language research.

A fundamental concern over the identified set is the exact delimitation of the research areas. Empirical evidence from the previous sections as well as background knowledge would indicate a reasonable distinction between the two research areas. However, no clearcut boundary can be drawn and, thus, the attribution of certain institutions to a particular area is problematic (such is LIMSI in France and the University of Nijmegen in the Netherlands).

Another empirical result reinforcing this point was finding 267 scientific publications that appeared in both research areas sets (speech and language). From this more limited bibliometric data set, a total of seven "collaboration groups" were identified, including AT&T Bell Labs, the University of Pittsburg, the MIT Artificial Intelligence Lab and the Stanford University Medical Informatics Section. It seems that no clear boundary should be drawn between the two research areas, particularly in applied research subspecialities. This also indicates some possible avenues for cross-disciplinary research, which is in accordance with the empirical evidence of later chapters in this thesis, particularly on the formation of a European Speech and Language research network.

A final topic that will be discussed briefly in this section is the importance of "expert" ex-post assessment of the "collaboration maps". In fact, some problems referred to above, and particularly in the language research area, relate to the inclusion in the research community of an extensive number of research groups working in research specialities such as Radiology, Psycholinguistics, Neuroscience and Medical Informatics. While their inclusion may indicate the significant interdisciplinary "breadth" of this research area it could also represent a lack of precision in the results. It is believed that only "expert" assessment would improve this ex-post screening process. This "expert" assessment is essential when using these "collaboration maps" for the purposes of research evaluation.

Nevertheless, the collaboration groups as defined by the co-authorship maps allowed the identification of a very significant proportion of the research institutions in the European area that are conducting research into Speech and Language. This is in accordance with the discussion in the following sections (see section 4.2), in particular regarding participation in research projects funded by the European Union in this research area, as well as "institutional" participation in the ELSnet network - European Network of Excellence in Speech and Language Research. Of particular note is the identification of active extra-European groups researching Speech and Language. These highly visible international research centres in Speech and Language were also identified from information on the individual collaboration analysed in the survey of the ELSnet researchers (the results of which are discussed in section 4.2.5).

The next subsection summarises the main findings of this section.

4.1.8 Conclusions and Further Research

The formal system of scientific communication, as represented by publication in refereed science journals, does indeed represent just the "tip of the iceberg" of the wide spectrum of formal and informal activities comprising the process of "scientific communication". However, this "tip of the iceberg" in itself constitutes an important source of evidence on the practices of communication and collaboration within research areas. This section has analysed empirical evidence on the speech and language research field, based on a bibliometric analysis of the publications indexed by the ISI - Science Citation Index electronic bibliographic database. The co-authorship analysis of bibliometric data for the period 1981 - 2001 provided a basis for the main theoretical discussions about the structure of scientific communication and the identification of "collaboration structures" and patterns of collaboration within this inter-disciplinary research area. Moreover, a longitudinal analysis of co-authorship activities over a significant time-period (1981 – 2000), provided good evidence for analysis of the evolution of this reseach field. Nevertheless, at the end of the analysis, more robust results were achieved by complementing this bibliometric analysis with other longitudinal analyses of collaboration (e.g. in research projects) as well as through a direct survey of active scientists in the field, who would self-report their communication and collaboration practices.

A first topic discussed in the chapter was the methodological problem of determining the "boundaries" of inter-disciplinary fields of research. On the assumption that no single optimal solution is available, this investigation has suggested a new methodology for delimiting the bibliometric data set, based on an Extended Broad Search method, combining Keyword Search, Core Journal Search and Core Author search techniques. The procedure seems to be robust and provided significant results in terms of coverage of co-authorships.

Secondly, the study has examined the relevance of "revolutionary contributors" in terms of their effects in structuring and determining the patterns of connectivity of collaboration networks. The empirical evidence on Automatic Speech Recognition seems to validate the hypothesis that "The winners structure it all". The structural impact was measured from multiple perspectives, including citation indicators, the impact on the "knowledge-maps" of the field, the extensiveness of immediately collaborating networks and the structural "citation impact" on the whole subfield of speech research and technology.

Collaborating networks are dynamic systems that continuously evolve over time. This research has provided evidence that co-authorship analysis can be used as an effective tool for mapping the structural dynamics of collaboration networks. The combination of a continuum approach and a long-term cross-sectional approach provides a mechanism for analysing the structural evolution of research fields. Some regularities in this evolutionary process were determined, such as the connectivity of the whole network, the "speciation" model of evolution and the emergence of a "core-periphery" structure.

Finally, the last section discussed the identification and characteristics of research "collaboration groups" based on the co-authorship results. Some important limitations of the methodology were highlighted and the main European nodes of the Speech and Language research network were identified.

A complete appreciation of the structure of research collaboration is probably beyond the scope of a single bibliometric analysis, no matter how effective. Nevertheless, these methods have provided a very good approximation of those structures. The combination of these patterns with information on collaboration, provided by a survey of the researchers in the community (section 4.2.4) and empirical results on research collaboration in European networks (section 4.2.3), provided evidence for mapping the "knowledge collaboration space" of European Speech and Language research.

Section 4.2 complements the bibliometric characterisation of research collaboration structures by discussing the "patterns" emerging from collaborative research projects funded by the European Comission in the field of speech and language. The emergence of a formalised "research network" is empirically tested. Following from the longitudinal analysis of research group collaboration in research projects (discussed in section 4.2.3), the information gathered from the survey of researchers active in this field provides good evidence of patterns of communication and collaboration (these personal collaboration patterns are discussed in section 4.2.4).

4.2 Research Networks as "Social Structures" for Scientific Collaboration and Communication

4.2.1 Introduction

In the previous section, we focused attention on the analysis of quantitative measures of scientific communication and particularly on indicators of the formal communication system of science. The discussion was based on bibliometric analysis and subsequent network analysis of co-authorship data for the speech and language scientific communities. The empirical evidence reinforced previous research results claiming that "co-authorship" indicators should be treated cautiously as partial and incomplete indicators of research collaboration (see e.g. Katz and Martin, 1997).

In this section, we will consider further empirical evidence on "research collaboration" structures, covering a wide spectrum of research "forms", from the informal and interpersonal "invisible colleges" to the more or less institutionalised research networks. The theoretical framework guiding the empirical analysis is organised in the following way. First, the relevance of research networks is positioned as one of a broad range of "forms" of collaboration and organisation within scientific communities. Secondly, we discuss the concept of "research networks" in the context of structures of collaboration and, particularly, as an outcome of deliberate science policy intervention. Finally, we emphasise the importance of electronic communication as an essential infrastructure leveraging the "networking" of research resources and collaboration. The empirical evidence comes from a longitudinal network analysis of European Commission (EC) research funding for the field of Speech and Language research, during the period 1990 - 2001, as well as from a survey of 312 researchers in the ELSnet network - the European Network of Excellence in Speech and Language Research.

The operational research question underlying the discussion is the identification of "forms" of research collaboration, for instance individual research networks at one extreme, and more formalised institutional research networks at the other, functioning as "social structures" influencing the process of electronic scientific communication. We attempt to gain a better understanding of this problem by analysing the following research issues. First, institutional "research networks" are dynamic systems, changing continuously over time, whose boundaries are fuzzy and not always clearly identifiable. This arises mainly from the multiplicity of relationships within a network, but also to a non-precise usage of the concept of a "research network". Second, there are, at the same time, persistent "structural" long-term collaboration linkages, detectable through longitudinal network analysis, generating knowledge that is highly relevant for science policy action. Third, the position a certain research institution occupies within a network, in terms of its centrality and connectivity, is important for understanding the organisation of the community as well as for the identification of hierarchical structures. However, this does not necessarily mean that the best-connected and most central research institutions are the leading "centres of excellence". Fourth, "institutional research networks" are not necessarily coincident with "individual research networks" in terms of composition, extension, trans-disciplinarity or other criteria. In fact, results from the longitudinal network analysis of collaboration in research projects, provide a first but incomplete approximation of the actual social structures embedded in communicative and collaborative processes. Information on the collaborative practices of researchers (gathered from the survey administered to ELSnet researchers) constitutes additional evidence enlightening the "hidden" structures of collaboration detected through the longitudinal analsysis. Fifth, the composition of networks, in terms of individual researchers' backgrounds and "seniority", may affect the pattern of collaboration within the network. It is in the light of these empirical findings that we analyse the relevance of electronic networking as digital structures supporting scientific communication and collaboration.

The section is organised as follows. In section 4.2.2 the relevant literature focusing on "networking" and collaboration is summarised and the operational research question and hypotheses in this section are situated in the context of a discussion of the concept of a "research network". Section 4.2.3 discusses the results of the longitudinal network analysis of EC funding to this research area and identifies long-term, well connected structures of institutional collaboration. Following from the longitudinal analysis of institutional collaboration, in section 4.2.4 we start the analysis of patterns of collaboration and connectivity at the level of researchers, based on the electronic survey of 312 researchers from the ELSnet network. The following topics are discussed in sequence in sections 4.2.4.1 to 4.2.4.5: international and inter-sectoral structures of collaboration; interdisciplinarity of research as reflected by researchers' background; collaboration patterns and "seniority" in terms of research career; individual researcher's collaboration networks; and the interdisciplinarity of collaboration (both for research groups and individuals). Section 4.2.5 explores the relevance of electronic networks as infrastructures supporting scientific collaboration and communication. The conclusions and further areas of investigation to be covered in later sections are summarised in section 4.2.6.

4.2.2 Funded Research Networks as "Institutionalised" Research Networks

Science policy has been reasonably active in endorsing the creation and development of "research networks". Some of the motives and incentives have been discussed in previous sections. It should be stressed that this is not a phenomenon limited to European research policy, but applies to other world regions. In the case of Japan, efforts date back to 1996 with the explicit funding of networks of excellence (via the State Promotion Programme aiming at establishing Centres of Excellence - CoE). At present, in Japan nine centres are funded with a total of around 36 million Euros. These research centres work in well-defined areas, such as "Global Information Processing Technology" and the "Study of the Creation of New Materials based on Insect properties" (AECMA, 2001).

In Canada, there are currently 11 Networks of Centres of Excellence, seven having been created in 1989 and four in 1995. It is estimated that the current funding allocated to these networks is of the order of $36 million annually (Canada NCE, 1998). In 1998 three additional networks were selected for initial funding. These networks span a wide variety of research topics from the Canadian Arthritis Network (a collaborative effort of more than 100 leading basic and clinical scientists from over 40 different institutions, with support from 39 companies); the Geomatics for Informed Decision (a new discipline, supporting a $10 billion industry around the world); and the Mathematics of Information Technology and Complex Systems (involving a partnership of about 70 industrial, financial and medical organisations, as well as around 400 researchers from 22 Canadian Universities) - (see Canada NCE, 1998).

In Europe, the networking efforts of the European Commission in terms of research and development began with the First Framework Programme, in 1984. Explicit emphasis has been given to promoting and endorsing the creation of "research networks" since the Fifth Framework Programme, in 1998, and particularly from 2000 with launch of a project to create a European Research Area (ERA) (European Commission, 2000), which has as one of its major instruments the creation of Networks of Excellence. The Sixth Framework Programme (for 2002 - 2006) explicitly considers as one of its major instruments in terms of European research strategy, the support of Networks of Excellence in certain fields of research. The "networking" goal of such instruments is explicitly stated in the technical reports of the EC.

> Networks of Excellence are designed to **strengthen scientific and technological excellence** on a particular research topic by networking together at European level

> the **critical mass** of resources and expertise needed to provide European leadership and to be a world force in that topic. This expertise will be networked around a joint programme of activities aimed principally at creating a **progressive and durable integration** of the research capacities of the network partners while, of course, at the same time advancing knowledge on the topic (European Commission, 2002, page 3 - emphasis in original).

This instrument of Networks of Excellence is complementary to the ideas of the ERA, concerning the networking of centres of excellence and the creation of virtual centres of excellence, as well as better use of the potential offered by electronic networks.

Materialisation of the concept of a "centre of excellence", as well as its reliable identification, are subject to intense debate. Science policy initiatives supporting the networking "form" are likely to affect the process of research collaboration and electronic research collaboration.

First, increased funds are likely to be allocated to the digital infrastructure of electronic networking, to facilitate the evolution of experience in this area and the collection of more empirical evidence on the use of ICT for scientific work (see section 4.2.6 on electronic networking of research resources). Secondly, it is likely that more visibility will be given to research centres (and hence research groups and researchers) participating in these research networks in the selected research topics, which in turn is likely to improve the "measurement" of "collaboration structures", as joint research activities such as research projects produce their scientific outputs. Thirdly, at least during the public "funding period", more formalised collaboration structures are likely to emerge, which will again facilitate the measurement process and, potentially, might improve the emergence of long-term collaboration structures. Fourthly, the networking of researchers and research groups is likely to enhance the intensity of communication between researchers, which is also important for the analysis of the changing nature of informal scientific communication and whether these processes are correlated with scientific productivity.

The following points sum up some of the previous analyses. First, the concept of a "research network" and its different forms should be understood as one among several other kinds of organisation of scientific and technological communities. Second, networks constitute a form of research collaboration, somewhere in a two-dimensional spectrum of "levels of collaboration" - from the individual to the international, and levels of "institutionalisation or formalisation" - from the completely informal "invisible college" to the formal research consortium. Third, a wide variety of factors motivate collaborative research, among which science policy intervention is certainly relevant. Finally, the "measurement" of research networks, performance or collaboration structure, is difficult, in which, although the problems have been more or less identified, the solutions to them are far less clear.

As the previous literature review highlights, a number of research topics are open for investigation.

Operational Research Question and Hypotheses

Our current research question focuses on identifying collaboration structures within a particular scientific community and understanding how these collaboration and communication structures relate to the adoption of electronic means of collaboration. In this section we will focus on the analysis of collaboration structures. Sections 5.1, 5.2 and 5.3 in chapter 5 will directly analyse the electronic networking of scientific collaboration.

A number of research issues frame the mapping of collaboration structures.

a) "Research Networks" at the institutional level are dynamic systems, that change continuously over time, and whose boundaries are fuzzy and not always clearly identifiable;
b) There are persistent "long-term structural" collaboration linkages, detectable through longitudinal network analysis;
c) The position that a certain research institution occupies within a network is relevant to science policy in terms of its centrality and connectivity;
d) For a research institution or group to become a "centre of excellence", it is a necessary but not sufficient condition that it should be among the best-connected and most central research groups within the research network;
e) "Institutional research networks" are not always coincident with "individual research networks", in terms of composition, extension and trans-disciplinarity among other factors;
f) The composition of a network, in terms of the backgrounds of individual researchers, as well as the "seniority" of researchers, affects the patterns of collaboration within the network.

The empirical evidence comes from two different data sets, reflecting the collaboration activity within the computational Speech and Language field of research. In order to analyse the institutional patterns of collaboration, a longitudinal analysis of research funding was conducted over the period 1990 - 2001. These data are for the European Commission research funding of this area, covering the Framework Programmes from 1990 onwards. An indepth network analysis of this data set allowed the identification of network structures as well as measurement of the connectivity and centrality of certain research institutions.

Analysis of collaboration patterns at the individual and research group levels was based on the results of an electronic survey of the set of researchers participating in the ELSnet network. The survey allowed the statistical analysis of 312 extensive questionnaires (corresponding to a 21.7% response rate), assessing the disciplinary backgrounds of individual researchers, their collaboration behaviour and their patterns of ICT use for scientific work.

The empirical analysis of the ELSnet network is significant for various reasons:

- at the European level it covers a wide spectrum of researchers, research groups and institutions dedicated to the topic of speech and language research;
- It is indeed a "Network of Excellence", which dates back to 1994, therefore the collaboration effects of its functioning should be "visible" now;
- It is an inter-disciplinary and inter-sectoral environment for research collaboration and communication;
- Given the absence of a research directory of the active researchers in this field, and the trans-disciplinary nature of the field of research - which represents serious obstacles to the use of bibliometric approaches to delimit the research field - this Network constitutes a representative roster of the researchers currently working on this topic.

The discussion of the results follows in this order. In section 4.2.3 we analyse the institutional patterns of collaboration resulting from the longitudinal analysis of research funding. The connectivity and centrality of some research institutions over the period 1990 - 2001 is also highlighted in this section. Following from the identification of "hidden social structures" of collaboration, as identified by the longitudinal analysis, the next section examines individual researchers' collaborative behaviour, using survey techniques. Section 4.2.4 analyses extensively individual patterns of collaboration, as revealed by the survey of ELSnet researchers. Finally, in section 4.2.5, we explore the topic of electronic networking of research collaboration, leading to a detailed examination of the use Internet technologies for research communication and collaboration (in sections 5.1 to 5.3).

4.2.3 Patterns of Institutional Connectivity (Analysis of European Funding)

In this section, we focus on the identification and characterisation of research collaboration structures at the institutional level within the European Speech and Language community. We use as a "proxy" for collaboration activity the participation of research institutions in co-operative research and technological development projects, funded by the European Community, during the period 1990 - 2001, under the Third, Fourth and Fifth Framework Programmes. Notwithstanding the fact that European funding of R&TD constitutes a relatively minor percentage (less than 5%) of the total amount of public funding of research across Europe[30] (Widhalm *et al.*, 2001), the structural impact of European research policies has been widely recognised, particularly in terms of its networking effects (which is our main focus here).

First, we briefly summarise European science policy in this regard and then discuss the empirical data testing the theoretical hypotheses defined in the previous section. In addition to the factors motivating research policy intervention in terms of networking of research, there are a number of specific reasons for the efforts in Europe on the topic of computational Speech and Language or Human Language Technologies,[31] First, there is the strategic research objective of pooling European human language technology research in order to compete (or reduce the gap) with the US in this field. Secondly, the language diversity of European countries constitutes a natural endowment, providing not only a need for new language techniques and methods, but also a very good potential for world excellence in "language engineering" research, which could overtake the traditional US advantage in speech research and technology. Thirdly, there is widespread confidence that the bridging of two traditionally separate communities (the speech research and technology community, on the one hand, and the language research and technology community, on the other) might result in a significant leap forward in the application of speech and language

[30] This note is taken from Widhalm *et al.*, 2001, page 139. "The EU spending on research accounted for about 4% in 1994 to 1998 (Research and Technological Development Activities of the European Union – 1999, European Commission (COM 99) 284, Tables 10 and 11, page 78f.). The FP represented only a very small share of total government spending on RTD in Member States and a very much smaller percentage of all RTD expenditures (public and private)".

[31] Given the inter-disciplinary nature of the field of research, the term "Speech and Language research" is used interchangeably with "Human Language Technologies research". The field of Human Language Technology covers a broad range of activities aimed at enabling people to communicate with machines using natural communication skills. Research and development activities include the coding, recognition, interpretation, translation and generation of language. The study of human language technology is a multidisciplinary enterprise, requiring expertise in linguistics, psychology, engineering and computer science (Cole *et al.*, 1997). It includes speech research and technology (automatic speech recognition, speech synthesis, speech prosody) and language engineering research and technology (computational linguistics, natural language understanding, language processing).

to innovation and society in general. Finally, there is a need for the integration of "centres of excellence" of European research in the field of speech and language (academic research centres, public research laboratories and private R&D laboratories) into "networks of innovation", which also involve users of these technologies in industry and society.

The above factors help to explain the continuous emphasis of European research policy on the field of speech and language. Moreover, they throw some light on the innovation character of the collaboration networks resulting from participation in pre-competitive co-operative RTD projects. Table 4-2-I summarises the major research programmes funded by the EC, in the field of speech and language.

Table 4-2-I - Summary of European Commission funding of Speech and Language Research (1990 - 2001)

European Human Language Technology Research and Technological Development Projects					
			Number of Institutions		
	Period	Nr. Projects	Univ.+ Public *	Private	Funding
Third Framework Programme	***- 1994***	49	81	93	**35 Million ECU**
LRE/TAP	- 1994				
MLAP	- 1994				
Fourth Framework Programme	***1994 - 1998***				**80 Million ECU**
Language Engineering	1994 - 1998	53	140	173	
MLIS	1996 - 1999	33	53	121	
ESPRIT	1994 - 1998	7	14	18	
INCO	1994 - 1998	12	53	37	
Fifth Framework Programme	***1998 - 2003***				**564 Million EURO ****
Human Language Technologies	1998 - 2002	50	144	114	
eContent	2001 -	7	7	23	

Notes:
LRE/TAP - Linguistics Research and Engineering / Telematics Applications Program
MLAP - MultiLingual Actions Plan
MLIS - MultiLingual Information Society Programme
ESPRIT - Long Term Research, selected projects in Speech and Language
INCO - International Cooperation Programme
HLT and eContent - both Programmes are integrated in Multimedia Contents and Tools Activity
Key Action III of IST Programme (Information Society Technologies Programme)
* This category includes both Academic Research Institutions and Public Laboratories. Other types of institutions were categorised as Private
** This is the total budget for Key Action III, including other projects other than Speech and Language

The above table confirms the significant investment made in European science policy, both in terms of the funds made available to this field of research and in terms of long-term and continuous commitment to supporting research in speech and language. In addition, network analysis of each of these research programmes reveals the international character of the collaboration, as well as the inter-sectoral nature of the research. This is to be expected, as the selection criteria for these co-operative research projects require the participation of at least three institutions from three different countries, as well as emphasises science-technology linkages of research activities. Research and development activities within these projects are conducted within a network of academic research centres, public research laboratories and private R&D laboratories.

It is important to note that our main concern in this research is the "networking" effect and, particularly, the identification and characterisation of collaboration structures emerging from these co-operative projects. From this perspective, the network analysis of these data sets produces interesting results. First, these "research networks" are dynamic systems, changing continuously over time. Secondly, and notwithstanding the above, some of these collaboration linkages are persistent, and give rise to long-term collaboration structures. Thirdly, the prolonged persistence and position of certain research institutions within these networks is significant, particularly in terms of their "centrality" and "connectivity". Finally, a central position within the network is a necessary, but not a sufficient, condition for such research institutions to become "centres of excellence". We turn now to the discussion of these results.

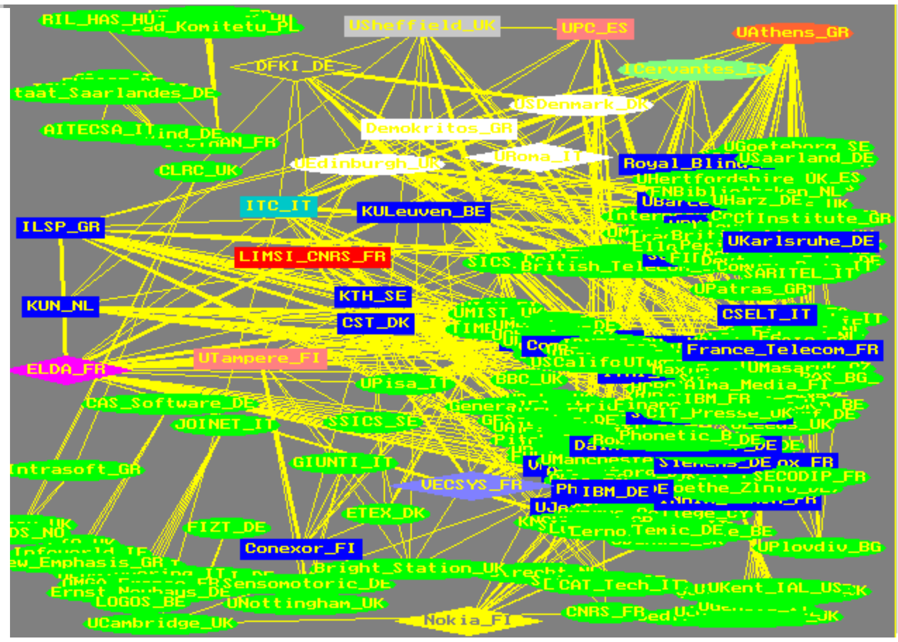
Each research programme involves a particular set of participating research institutions. Each institution participates in one or more research projects, in collaboration with other research institutions. The combination of the whole set of institutions and all the linkages between them (a link being mutual participation in the same research project) forms a "research network". For the analysis here, we have focused on seven different "research networks". The longitudinal analysis reveals that the composition and structure of these networks is likely to vary significantly over time, but also apparent is the existence of some long-term "patterns" of collaboration, reflecting a kind of "hidden social structure" supporting collaborative practices.

These different "research networks" should be treated as different structural entities. The number of research institutions varies significantly, the research institutions participate in different ways in different networks and the nature of the collaboration links is structurally different in terms of the connectivity of the whole network and the connectivity of each individual institution.

For example, the "research network" formed for the Human Language Technologies programme, running over the period 1998 - 2002 and involving 258 different institutions collaborating on 50 research projects, is structurally different from the "research network" formed for the Multilingual Information Society Programme, funded over the period 1996 - 1999, and involving 174 institutions in 33 research projects. Figures 4-2-1 and 4-2-2 depict the structure of these two networks.

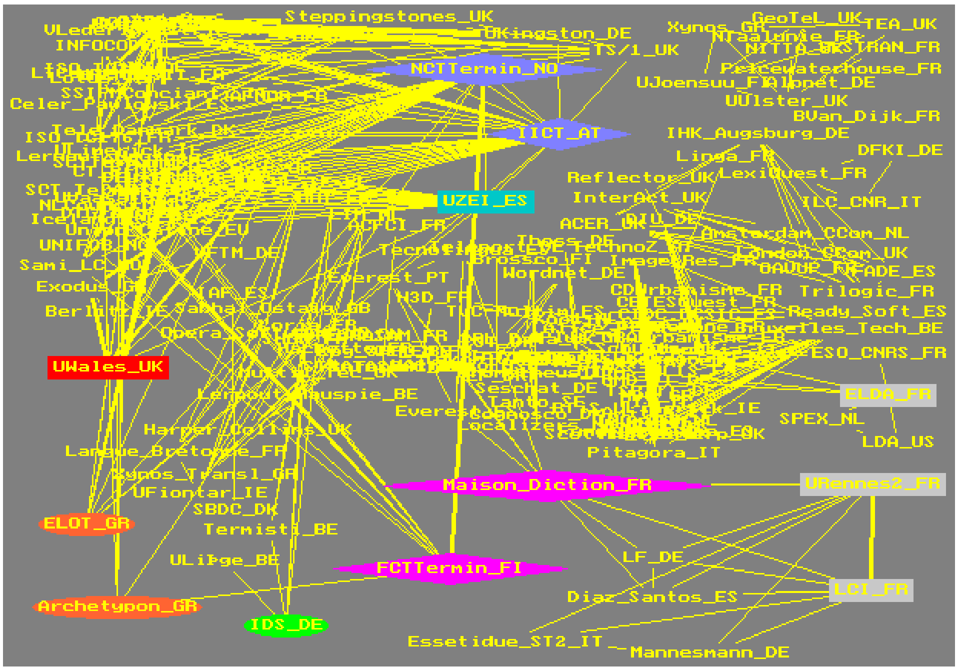
The graphical representation of these collaboration networks reveals the apparent complexity of the interactions among various research groups. However, it is also apparent from the diversity of linkages and connectivity patterns resulting from the collaborative activity, that there is a "hidden structure", which explains the persistence of collaboration over extended periods of time. A better characterisation of these "hidden collaboration patterns" can be obtained using other techniques such as interrogating individual participants in these collaborating groups (see results of survey in section 4.2.4).

Figure 4-2-1 - "Research Network" Human Language Technologies Programme (1998 - 2002)



It should be stressed that the graphical position of any one institution within these network maps is not relevant. They are only a visual representation and the same network structure could be represented in several different ways. What is important is the number and composition of the network members (the institutions participating in the network) and the number and type of links connecting each institution with other institutions. For instance, it is of relevance that the University of Sheffield is connected with the Universidad Polytecnica Catalunha in Spain, with DFKI in Germany, and with the KUN research centre in the Netherlands. In addition, DFKI is connected with the University of Sheffield, the KUN research centre, LIMSI in France, the Instituto Linguistica Computazionale in Italy, the University of Edinburgh in the UK and the University of Southern Denmark. On the other hand, it is completely irrelevant that both DFKI and the University of Sheffield are represented at the top of this network map, other than for practical visualisation reasons.

Figure 4-2-2 - "Research Network" Multilingual Information Society Programme (1996 – 1999)



Comparative analysis of the two research networks highlights certain structural differences between them. What is important is that the empirical evidence appears to confirm the theoretical hypothesis that research networks change continuously over time and, as such, are dynamic systems.

If we assume a complementary but different point of view, we might be analytically and methodologically more interested in considering the whole research community of Speech and Language as the "Research Network". In this sense, we should treat this community of practitioners as the "Research Network of Speech and Language".

From this perspective, we now have a single "research network", constituted by the whole set of research institutions participating in any of these research programmes over the period 1990 - 2002. If this were the case, we need to take into account that: 1) we have changed the concept of "research network"; and 2) we now have *multiplexity*[32] in our analysis. This means that a single group of entities (the whole set of research institutions) participates in different "relations" or individual networks. Each relation is given by the participation in a different research programme over time. Thus, we have a "single research network" that is represented by seven individual "social networks". In this new

[32] *Multiplexity* has the technical meaning that the same set of members is measured in different relationships, resulting in different networks.

methodological framework, we have the additional advantage of being able to analyse the evolution of structures of collaboration, or at least to grasp the "hidden structure" of collaboration that can be better explained by complementary survey techniques.

This longitudinal network analysis provokes two important comments. First, some research programmes or networks are more structurally similar than others. Secondly, there are some collaboration patterns that persist over time, i.e. some institutions participate often in several different networks and, moreover, collaborate in a "core" group over time.

The structural analysis reveals two separate and distinguishable groups of programmes. A first group of research programmes is specifically focused on speech and language engineering. This group includes the research programmes in Framework Programme 3 and before that (in a period prior to 1994), the Language Engineering Programme (1994 - 1998), and the Human Language Technologies Programme (1998 - 2002). A second group, with a more diversified set of research objectives and not directly oriented to speech and language, includes the research projects conducted under the ESPRIT programme (1994 - 1998), the Multilingual Information Society Programme (1996 - 1999), the INCO International Co-operation Programme (1994 - 1998) and the E-Content Programme (2001). These research programmes were oriented towards the production and dissemination of multimedia content, translation services and the application of multilingual services in the information society.

The type of content of the research programmes is clearly reflected in the composition and structure of their corresponding networks. Their differences are significant in terms of participating institutions, connectivity of the networks and the collaboration groups formed within these networks. At the same time, a closer analysis of the persistence of collaboration groups within these networks, and particularly an analysis focusing on the three networks dedicated to speech and language research, appears to confirm the continued presence of certain institutions as well as a long-term collaboration pattern. The significance of these "hidden social structures" of communicative and collaborative activities could only be fully ascertained by collecting information, for instance, via a survey, about researchers' collaboration (see discussion in section 4.2.4).

Table 4-2-II summarises those research institutions that participated actively in all these research programmes and are persistent members of these networks.

Table 4-2-II - Group of Research Institutions with Long-term Participation in Speech and Language Networks.

Long-term active Institutions in Speech and Language networks
CLI - Computational Linguistics Institute - Italy
CPK - University of Aalborg - Denmark
CST - Centre for Speech Technology - Denmark
DFKI - German Research Centre for Artificial Intelligence - Germany
ILSP - Institute for Language and Speech Processing - Greece
INESC - Portugal
LIMSI - CNRS - France
KUN - Netherlands
KTH - Sweden
KULeuven - Belgium
IGFAI - Institute for Artificial Intelligence - Germany
UPC - Polythecnique University of Catalunia - Spain
University of Stuttgart - IMS - Germany
University of Sheffield - United Kingdom
University of Utrecht - Nertherlands
University of Edinburgh - United Kingdom
Rank-Xerox - France

There is no relevance to the ordering of these institutions. The important point that emerged from the evidence is that all of them are among a "core" group of institutions collaborating often within these networks over a significant period of time (1990 - 2002) – and also are members of the ELSnet network. This long-term structural collaboration leads us to a discussion of the relevance of the structural position of a research institution within these networks. We examine this topic in the following paragraphs.

In order to assess the centrality of each research institution within these networks, two complementary indicators are relevant in terms of structural analysis. The first indicator - *degree of centrality* - measures the "activity" of each institution within the network and is based on a count of the total number of collaboration links with other research institutions. A second, and complementary indicator - *betweenness centrality* or *degree of connectedness* - measures the importance of each research institution to the connectedness of the whole network, i.e. the bridging function of one institution in linking separate "groups" making up the whole network.

The empirical evidence based on the analysis of centrality of institutions, over the period 1990 - 2002, confirms the relevance of the structural position of research institutions, namely the level of degree of centrality and connectedness. Some comments should be made.

The degree of centrality indicator corroborates the previous results that a smaller set of institutions is persistently more active within these networks. The network analysis over time allows one to select the group of "most active institutions". Table 4-2-III shows the results for the top percentile of most active institutions within the network of Framework Programme 3 and before.

Table 4-2-III - More Active Research Institutions within Network Framework Programme 3 and Before (- 1994)

Human Language Technologies - Framework Programme 3 and Before

Centrality Degree - Measure of Activity within Network		
ICL_CNR_IT	56.000	32.370
UStuttgart_IMS_DE	41.000	23.699
CST_DK	34.000	19.653
UUtrecht_STT_NL	32.000	18.497
ILTEC_PT	24.000	13.873
UEdinburgh_UK	23.000	13.295
IAI_DE	22.000	12.717
UMIST_UK	22.000	12.717
KULeuven_BE	22.000	12.717
TALANA_FR	21.000	12.139
UEssex_UK	21.000	12.139
SITE_EUROLANG_FR	21.000	12.139
Lingsoft_FI	20.000	11.561
UAmsterdam_NL	20.000	11.561
LIMSI_CNRS_FR	18.000	10.405
ILSP_GR	18.000	10.405
UPompeu_ES	18.000	10.405
Dima_IT	18.000	10.405
INESC_PT	18.000	10.405

Note: The first column represents a first computation of the degree of centrality measure. The second column is the final measure for this indicator. It should be stressed that the absolute value of the indicator is not relevant - what is important is the relative value among the set of members of the network.

An additional empirical finding is that not only do the most active institutions participate more often in various networks, but they also collaborate more with other active institutions. This generates some degree of "clustering" within the whole network, as well as leading to "zones" of the network revealing higher levels of connectedness, compared with other more sparse collaboration zones.

If the centrality degree allows one to select the "core" of most active institutions over time, a more refined structural indicator is given by the inter-linking role of certain institutions, connecting several groups within the larger network. This is measured by the *betweenness* indicator. Table 4-2-IV presents the results for the Human Languages Technology Programme network.

Table 4-2-IV - "Best-Connected" Research Institutions within Network HLT (1998 - 2002)

Human Language Technologies - 1998 - 2002

Betweenness - Measure of Connectedness within Network		
DFKI_DE	6.708.027	20.392
UAthens_GR	5.367.767	16.317
ITC_IT	4.037.559	12.274
UEdinburgh_UK	3.174.200	9.649
USheffield_UK	2.975.892	9.046
ILSP_GR	2.510.491	7.632
KUN_NL	2.305.121	7.007
UGesamt_DE	2.120.680	6.447
USDenmark_DK	2.062.217	6.269
LIMSI_CNRS_FR	1.871.015	5.688
CST_DK	1.831.438	5.567
ELDA_FR	1.796.481	5.461
France_Telecom_FR	1.784.089	5.423
INRIA_LORIA_FR	1.540.000	4.681
Telefonica_ES	1.414.772	4.301
KTH_SE	1.254.330	3.813
UPC_ES	1.243.645	3.781
UTampere_FI	1.243.527	3.780
RankXerox_FR	1.110.000	3.374
Sony_DE	977.717	2.972
KULeuven_BE	957.647	2.911
ITRI_UK	953.792	2.899
ICervantes_ES	930.270	2.828
UPM_ES	928.136	2.821
Nokia_FI	853.361	2.594

Note: The first column represents a first computation of the *betweenness centrality* measure. The second column is the final measure for this indicator. It should be stressed that the absolute value of the indicator is not relevant - what is important is the relative value among the set of all the members of the network.

These institutions are quite important to the structural connectedness or "integration" of the network. Their importance can be seen in the special role they assume in collaborating more intensively with different parts of the network, allowing the exchange of information and the integration of collaboration structures. Some empirical results are worth commenting on. First, some of the most active institutions are also the "best-connected ones". This is confirmed over time. The former is a necessary but not a sufficient condition for the latter. Secondly, the results confirm the theoretical hypothesis that the position an institution occupies within these networks is highly relevant, particularly in terms of integrating the collaborative efforts.

The final hypothesis regarding whether or not the most active and best-connected institutions are the leading "centres of excellence" can only be partly answered in structural terms in the light of the available empirical evidence. Assuming there is some consensus over the concept of a "centre of excellence",[33] the above results support the argument that to be very active and to be among the "best-connected" research institutions within a research network is a necessary, but not a sufficient condition, for becoming a centre of excellence.

The structural analysis indicates that the visibility of a research institution, over a reasonable period of time, is dependent on the activity of that institution within the network. Moreover, the capacity of one institution to participate in several different collaboration groups is also relevant for the integration of the whole network. We would expect that for one institution to become recognised as a "centre of excellence", it should be visible. The real conundrum comes with the meaning of "excellence". It is outside the scope of this investigation to measure scientific "excellence" at the institutional level – or, indeed, to decide whether that might be a feasible endeavour. However, if one assumes that a necessary condition for becoming "centres of excellence" is participation within collaboration networks, then we have structural evidence that confirms the above hypothesis: to be among the "centres of excellence" a research institution should be active and very well connected. A further methodological analysis along this line of investigation could be through expert "peer review" of these active and well connected institutions, and their inner collaborations, in order to assess their "excellence".

In this section we have analysed patterns of research collaboration at the institutional level. The empirical evidence of a decade of funding of collaborative R&D projects provided the data for a discussion and confirmation of four key hypotheses. Research networks need to be rigorously defined in terms of composition and structure as they are indeed dynamic systems, changing continuously over time. Secondly, the long-term analysis of network dynamics allows one to identify groups of collaborating institutions forming persistent patterns of research collaboration. Thirdly, we found evidence to confirm that the centrality of research institutions within networks, in terms of activity and connectivity, is highly relevant for the "integration" of these research networks. Finally, we partially explored the idea that to become a "centre of excellence" a research institution should be active and very well connected within these networks. This would lead us to believe that these institutions have a significant role with regard to the intensity of communication within the whole scientific network.

---

[33] The Norwegian Ministry (2000, p. 2)offers a definition of a "Centre of Excellence": "The CREST working group has defined a centre of excellence as "… centre of cutting edge research in emerging fields with emphasis on innovation".

As noted in the literature review section, there is a wide spectrum of collaboration "forms", going well beyond "institutional" collaboration structures and "research networks" resulting from science policy. As discussed in this section, these institutional collaboration structures are highly relevant for identification of the key-players within a field of research. However, some of the most important collaboration linkages are hidden within informal and individual practices. In the next section, we analyse more deeply the collaboration and communication practices of individual researchers. The empirical evidence comes from a survey of 312 researchers belonging to the European Network of Speech and Language. We are particularly interested in understanding individual collaboration networks, the nature of the collaboration - in terms of inter-disciplinarity, international and inter-sectoral collaboration - and the differences in collaboration patterns arising from different levels of seniority in research.

4.2.4 Patterns of Collaboration and Connectivity at the Researchers Level (Analysis of Survey Results)

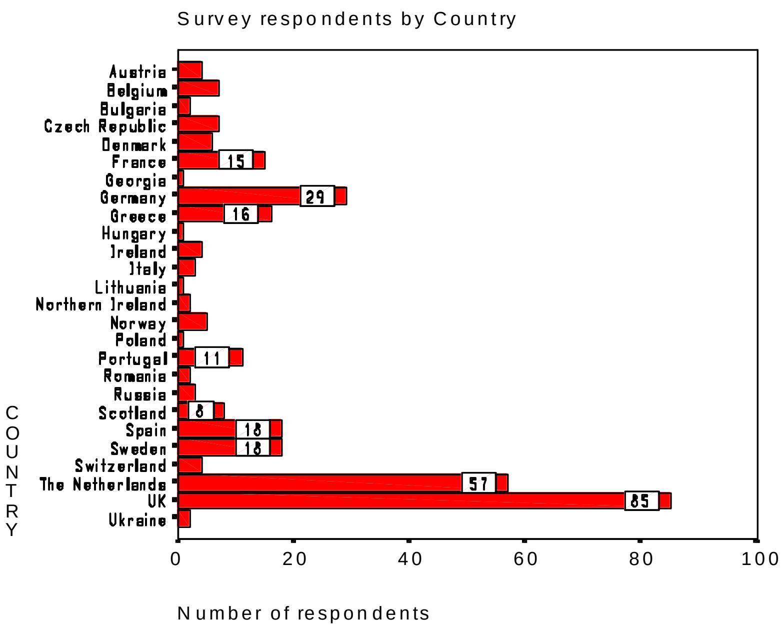
In this section we analyse, at the level of the individual researcher, the patterns of collaboration in collaboration networks and communication practices. This analysis brings more empirical evidence on the nature of collaboration activity, previously identified by the longitudinal analysis of communication and collaboration, whether through publication or participation in cooperative research projects. The empirical data come from a survey of 312 researchers working in the field of computer Speech and Language, belonging to the ELSnet network. The survey collected information about individual scientists' disciplinary backgrounds, their collaboration practices and their use of ICTs for collaboration and communication. Sections 4.2.4.1 to 4.2.4.5 discuss the first two aspects. The use of ICTs for research work is discussed in chapter 5, section 5.1.

This section is organised as follows. First, we discuss the international and inter-sectoral nature of collaboration, as reflected by the collaboration patterns of the survey respondents. Second, we analyse the wide diversity of disciplinary backgrounds characterising the researchers belonging to the speech and language communities. Thirdly, we find how collaboration patterns in terms of international and local practices might be associated with "seniority" of researchers - as given by their experience in terms of number of years of research. Fourth, we compare the differences between individual "invisible college" networks - as defined by the nominations of researchers of their closest collaborators - and institutional research networks, such as those analysed in the previous section. Finally, we analyse the significance of inter-disciplinarity in collaboration, in terms of the research groups with which each researcher works, as well as the more intimate individual collaborations.

4.2.4.1 International and Inter-sectoral Structures of Collaboration

The international nature of the research community as well as its inter-sectoral character has been noted in previous sections. At the level of the individual researcher, there is also a wide range of international contributors to the community. Figure 4-2-3 shows the breakdown by country of the survey respondents.

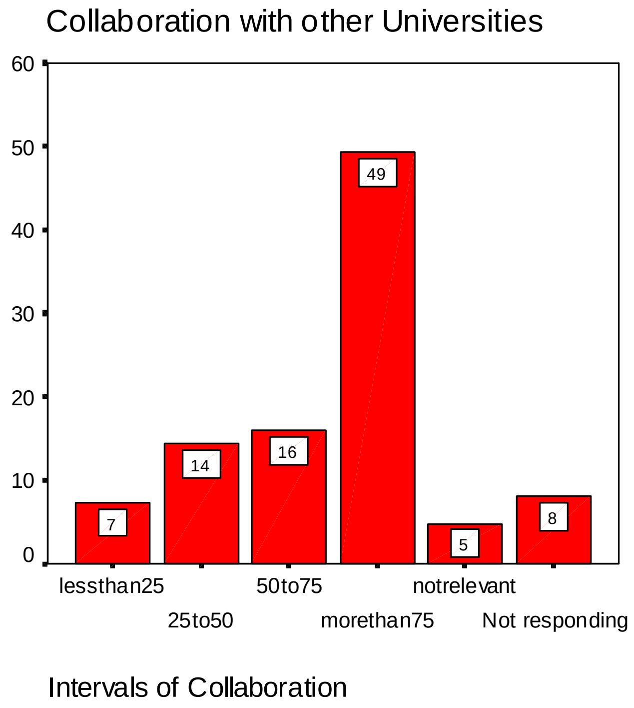
Figure 4-2-3 - Survey Respondents by Country



The total number of respondents by country is proportional to the total number of active researchers who are members of the research groups of ELSnet (correlation of 0.98). It can also be seen from this table that there is indeed a wide diversity of European countries contributing to research in computational speech and language.

Analysis of the collaboration between sectors at the individual researcher level (for example a university researcher collaborating with a public laboratory researcher or a private R&D laboratory researcher) does not emphasise the inter-sectoral nature of the community. However, it is important to remember that the majority of survey respondents belong to academic research departments and public laboratory research centres. As such, the inter-sectoral patterns of collaboration are likely to be under-estimated. Figure 4-2-4 depicts the results of the analysis of inter-sectoral patterns of collaboration, and particularly the importance of collaboration with university research groups.

Figure 4-2-4 - Collaboration with University Research Groups



The data appear to confirm the importance of intra-university research collaboration, as well as academic - public laboratories collaboration (but the graph only shows universities). Almost 50% of respondents who considered collaboration with other universities to be relevant, indicated that they collaborated with universities in 75% of cases. An additional group of 16% of respondents indicated that they collaborated in the range of 50% - 75% of cases with other research groups from universities.

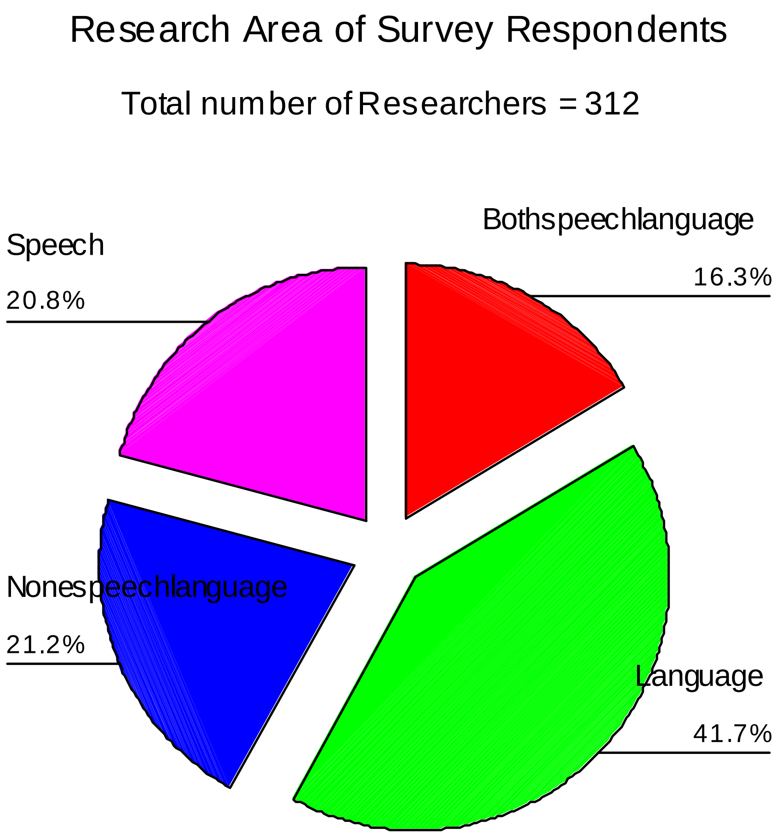
In the next section we turn to the analysis of the inter-disciplinary patterns of collaboration. Theoretically, we would expect a wide diversity of backgrounds contributing to such a diverse field of research.

4.2.4.2 Interdisciplinarity of Research as Given by Researchers' Background

When we analysed quantitative measures of scientific communication (in section 4.1) using bibliometric analysis of publication in scientific journals, we found a wide diversity of contributors to the "speech and language" research community. Indeed, we found that originally, and even nowadays, these two communities had their own separate sets of journals, conferences and other specialised communication fora. Therefore, it is no surprise to find that there is a wide variety of researchers contributing to the research community, and a variety of affiliations with research specialities within speech on the one hand, or within language, on the other. Hence, we would also expect to find a wide diversity of disciplinary backgrounds among researchers, given the inter-disciplinary character of the research field.

The empirical data collected at the level of the individual scientist confirms both hypotheses.

Figure 4-2-5 - Research Area of Individual Scientists Working in Speech and Language



While 41.7 % of the respondents consider themselves to be researching in the "Language" research area, 20.8% position themselves as working in "Speech" research. There is an overlapping region - "Speech and language" - applying to 16.3% of the respondents who considered themselves to be researching in both areas. These results auger well for the "integration" of the two sub-specialities within a single research community. The significant result that 21.2% of respondents considered themselves to be working neither in speech nor in language is explained by differences among individual scientists in their conception of the term "research area". In fact, this category of respondents chose more fine-grained research specialities as their "research area". Among the specialities indicated are linguistics, theoretical linguistics and computational linguistics, language processing, phonetics, speech processing, speech prosody, pattern recognition, information retrieval

and automatic translation. This is an indication of the degree of "division of labour" and specialisation within the community.

The inter-disciplinary nature of the community is also revealed from an examination of the disciplinary background of the researchers. Researchers had the opportunity to indicate, in a free-choice setting, the discipline in which they were granted their doctorate. From the total set of 171 researchers holding a doctorate, Table 4-2-V shows the breakdown by "discipline" of specialisation of 113. The full list of doctoral specialisations is given in Annex VIII.

Table 4-2-V - Disciplinary Background of Researchers Based on Doctorate

Disciplines of Doctorate with more frequency	
	Number of Researchers
Artificial Intelligence	10
Computational Linguistics	13
Computer Science	32
Electrical Engineering	5
Linguistics	33
Mathematics	7
Speech Science	13
	113

It is also noteworthy that there is a relatively well-defined differentiation pattern between speech researchers and language researchers. While the former are more likely to have a background in Speech Science or Computer Science, language researchers have a disciplinary specialisation in Linguistics or Computational Linguistics. Even if one allows for these differences, the empirical evidence corroborates the wide diversity of disciplinary backgrounds of researchers contributing to this field of research.

These results are in accordance with the analyses conducted in previous sections characterising research collaboration at the institutional level. In fact, among the more active and best-connected institutions within these networks, identified by network analysis, there are some institutions more specialised in speech research and others that focus on language research.

The next section tests the hypothesis of whether collaboration patterns change according to the researchers' "seniority" given by the number of years spent in research. The theoretical expectation that more experienced researchers may have higher levels of research productivity - in terms of number of papers and co-authored publications - is also examined.

4.2.4.3 Collaboration Patterns and "Seniority"

An interesting characteristic of the composition of research groups, research networks and research communities is the "seniority" of researchers as given by the number of years spent in research. This might be relevant not only in terms of total research productivity (at the various levels), but also as a factor determining different patterns of collaboration. We turn now to the discussion of both of these aspects.

Table 4-2-VI summarises the results on the seniority of the respondents to the survey.

Table 4-2-VI - Survey Respondents "Seniority" - Number of Years of Research.

yearsresearch

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid		2	.6	.6	.6
	lessthantwo	41	13.1	13.1	41
	uptofive	86	27.6	27.6	100.0
	fivetofifteen	117	37.5	37.5	38.1
	morethanfifteen	66	21.2	21.2	72.4
	Total	312	100.0	100.0	

From the total set of 310 valid responses, almost 59% have at least five years of experience in this research area and 21.2% have more than fifteen years.

A theoretical expectation is that more "senior" researchers will show better research performance. We use as a "proxy" for research performance or productivity, two complementary indicators - the total number of papers published in the last five years and the total number of co-authored papers published in the last year.

Table 4-2-VII - Comparison of research performance across different categories of “seniority”

Report

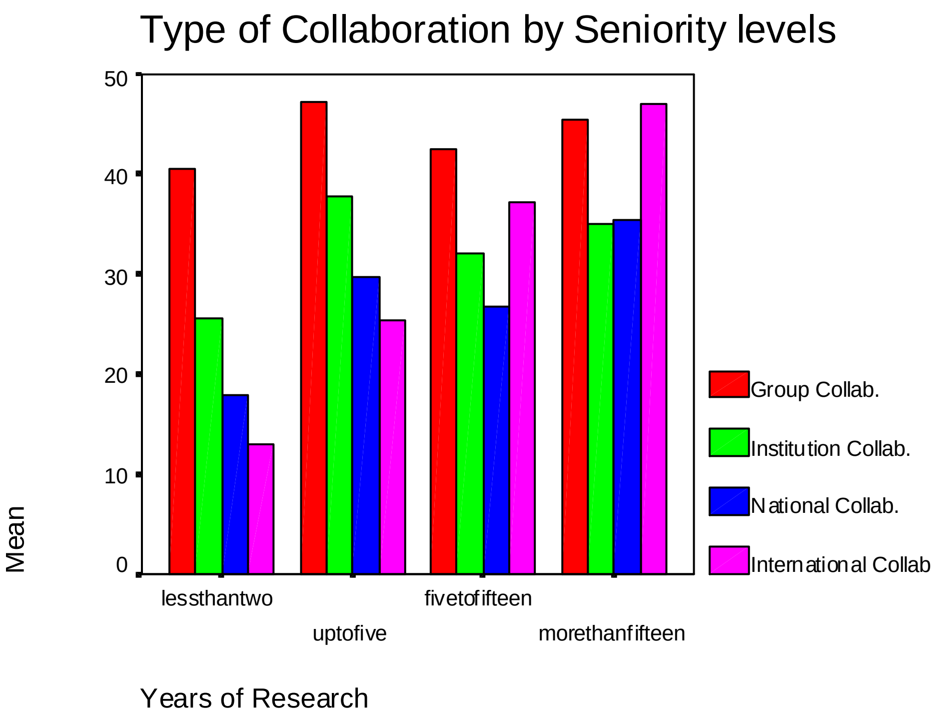
yearsresearch		nrcoauthor lastyear	totalpapers
lessthantwo	Mean	2.1481	1.8889
	N	27	9
	Std. Deviation	1.99429	.92796
	Minimum	.00	1.00
	Maximum	10.00	4.00
uptofive	Mean	2.4783	2.4865
	N	69	37
	Std. Deviation	1.62344	2.26840
	Minimum	.00	1.00
	Maximum	10.00	10.00
fivetofifteen	Mean	3.0737	4.3721
	N	95	86
	Std. Deviation	2.79529	4.14458
	Minimum	.00	1.00
	Maximum	22.00	27.00
morethanfifteen	Mean	4.7400	5.8077
	N	50	52
	Std. Deviation	5.57623	5.35793
	Minimum	.00	1.00
	Maximum	30.00	27.00
Total	Mean	3.1452	4.2772
	N	241	184
	Std. Deviation	3.37016	4.31598
	Minimum	.00	1.00
	Maximum	30.00	27.00

The empirical tests confirm that there are significant statistical differences among categories of “seniority” in terms of an individual researcher’s productivity. Moreover, research performance appears to increase with the number of years of research. More experienced researchers (more than fifteen years of research) show greater productivity than researchers with five to fifteen years experience and, in turn, this category of researchers produces better performance than the less experienced categories.

Also, this “research performance” is not affected by whether the research activity giving rise to a publication is individual or collaborative. In fact, the results are the same for co-authored publications, which also indicates that in relative terms more experienced researchers tend to collaborate more.

However, a particularly interesting empirical finding from this analysis is that “seniority” in research is not only associated with research performance and, potentially, intensity of collaboration, but also is statistically significantly correlated with different patterns of collaboration. Figure 4-2-6 summarises the results of this analysis.

Figure 4-2-6 - "Seniority" in Research and Different Patterns of Research Collaboration



When one compares different types of collaboration (collaboration within the institution, with other national research groups or with other international researchers and research groups) across categories of "seniority" in research , significant differences emerge. More experienced researchers, i.e. those in both the five to fifteen years and the more than fifteen years groups, are likely to give relatively more importance to international collaboration in the course of their research than the less experienced researchers. Contributory factors might be the accumulation of contacts to network with, as well as the greater "visibility" and prestige of these researchers at the international level. Less experienced researchers tend to collaborate within the boundaries of their particular research group or research institution.

4.2.4.4 Individual Researcher's Collaboration Networks

This section analyses empirical evidence on personal collaboration networks, and attempts to identify patterns of collaboration resulting from the communication practices of these individual researchers. The survey included a specific question about researchers' personal collaboration networks. They were requested to specify up to six research groups with whom they collaborate the most, in descending order of intensity. These data allowed a mapping of the structure of the collaboration networks for individual scientists.

A first assumption is that, at the level of individual scientists, there will be as many research networks as the total number of researchers under analysis. In fact, each personal network constitutes what might be called a personal "invisible college", composed of the closely knit and tightly bounded group of collaborators with whom the researcher most intensively conducts research. This assumption is empirically validated by the results. In fact, the whole network resulting from the aggregation of individual personal networks is a very sparse, highly disconnected network. This should come as no surprise. We would expect that individual researcher's networks would be much more sparse than the institutional networks examined in previous sections, given the likely greater variety of relationships conducted in individual interactions.

On the other hand, this increased variety in the number and type of collaborators also brings more diversity to the whole research network. Indeed, very well-connected scientists might link entire research groups to other "external" research groups. This is usually considered a good way of accessing new and more diverse research information and resources. For example, researchers from DFKI revealed intensive collaboration with the Computer Science Linguistics Institute of Stanford University, which is not part of the European research network in speech and language, but is definitely a world-leading research group in this field. Another similar example is the citing of ATR in Japan by several researchers belonging to different but closely collaborating research groups (the ICP Institute of Speech Communication and the LPL - Laboratory for Speech and Language CNRS, in France). A major conclusion from this empirical analysis is that the variety of personal research networks offers better potential for the discovery of knowledge resources, resulting from long-term and persistent personal interaction.

A second important result is that analysis of personal collaboration networks might reveal important patterns of collaboration at the institutional and research group level. The analysis of a reasonable number of personal networks from scientists belonging to the same research group might be a very good indication of the collaboration pattern of the entire research group.To exemplify this, Table 4-2-VIII presents a small part of the total collaboration matrix.

Table 4-2-VIII - Portion of Collaboration Matrix Resulting from Aggregation of Personal Research Networks

	UCL_UK	Ubirmingh	MaxPlack_	Ucambridg	CPK_DK	TeleDanMa	Usaarbrueck	SonyStutt
UCL_UK	6	5	4	3				
Ubirmingham_UK		1						
MaxPlack_NK			1					
Ucambridge_UK				36				
CPK_DK					1	6	5	
TeleDanMark_DK						1		
Usaarbruecken_DE							1	
SonyStuttGart_DE							5	1

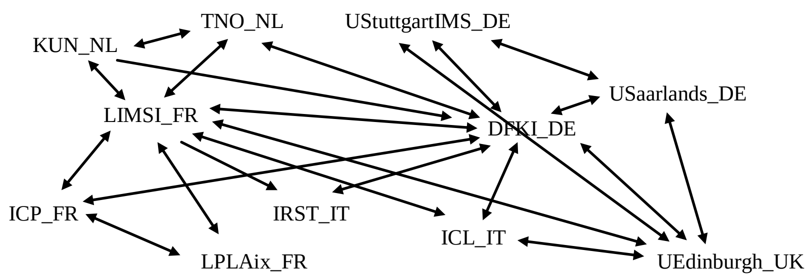
Note: Cell values indicate the strength of association between research institutions and results from the relative importance given by each respondent to the collaborating research groups.

The table reveals the intensity of *intra-institutional* collaboration for the University of Cambridge. Indeed, several researchers indicated other research groups within the same institution as well as the research group of which they were members, as being among the main and most important collaboration partners.

In addition to the measurement of intra-institutional collaboration, these personal networks also reveal other regularities in the patterns of *inter-institutional* collaboration. We are able to identify sets of research groups belonging to different institutions, and usually different countries, that collaborate more intensively with each other. It should be stressed that this represents a more accurate method of mapping research collaboration across different levels (individual, research group, departmental, and so on) than the traditional bibliometric methods, which constitute only partial and limited indicators of these patterns.

Figure 4-2-7 presents a graphical representation of the collaboration network, focusing on two of these zones of more intensive collaboration.

Figure 4-2-7 - Portion of Research Collaboration Network with Selected Core of Densely Connected Research Groups



This representation provides empirical evidence of more intensive collaboration patterns between DFKI in Germany and several other institutions (University of Saarlands - Computer Linguistics, Computer Science and Phonetics, University of Stuttgart-IMS, University of Edinburgh, TNO in the Netherlands, LIMSI in France and OEFAI in

Austria), as well as between LIMSI in France and also several other institutions (Laboratory for Speech and Language, Institute for Speech Communication in France, and IRST in Italy).

A comparison of these personal collaboration networks with the institutional research networks analysed in the previous section also corroborates the more intensive level of collaboration among a core group of the best-connected and most active institutions. This, then, provides empirical evidence of the visibility of those institutions and constitutes a reliable methodological tool for assessing the institutional collaboration patterns previously detected by the longitudinal analysis of institutional research funding.

Finally, a detailed analysis of some of the more densely connected collaboration groups also reveals important intra-disciplinary patterns. In fact, the major proportion of the more intensive collaboration relationships are with research groups working in the same speciality, which is in accordance with the results that will be discussed in the next section.

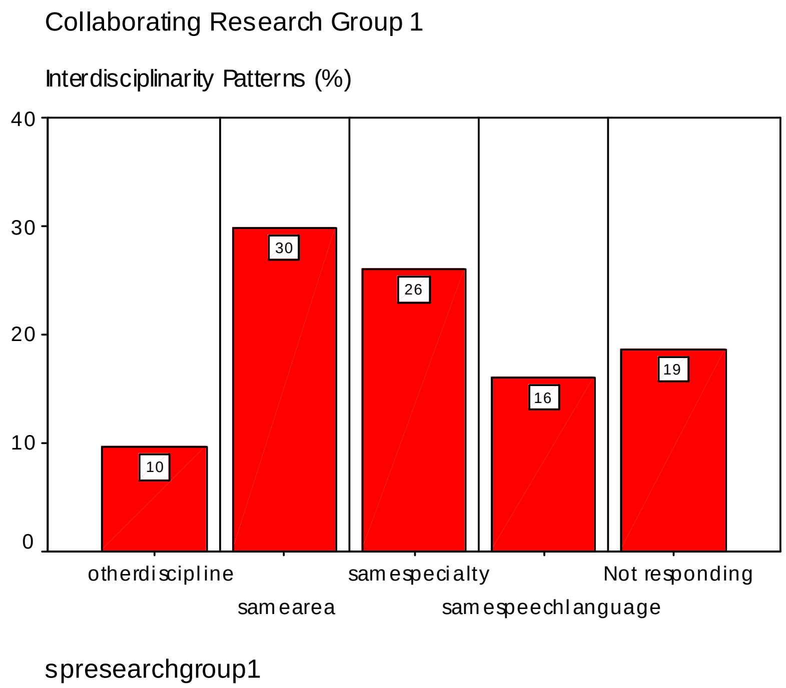
4.2.4.5 Interdisciplinarity of Collaboration (institutional and individual)

In this section we focus on the degree of "specialisation" of the collaboration at two complementary levels: the level of the research group and the level of the collaborating researchers.

Researchers were asked to indicate in the research area in which those they collaborated most closely with were working. Figure 4-2-8 shows the results.

Most respondents indicated that their most intensive collaborative activities are with research groups within the same research area (30%) or even within the same research speciality (26%). In addition, a further 16% of the respondents reported that research collaboration takes place in the field of speech and language. Only 10% of the respondents considered their most important collaborating research group to work in a different discipline. This is empirical evidence of a quite specialised collaboration pattern.

Figure 4-2-8 - Specialisation Pattern of More Intensive Collaborators



Note: Within the same research area, researchers were allowed to differentiate several research specialities.

This hypothesis is corroborated to a lesser degree from analysis of the specialisation pattern of the second and third most important collaborating research groups. Table 4-2-IX shows the results of the degree of specialisation of the collaboration with the second most important research group.

Table 4-2-IX - Specialisation of Second Research Group with Whom Researchers Collaborate

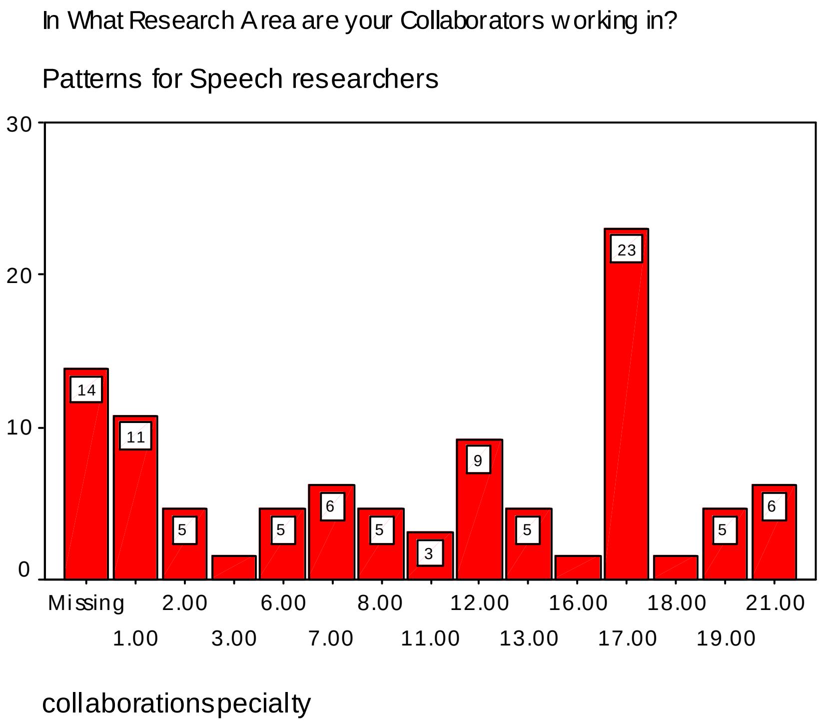
spresearchgroup2

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid		95	30.4	30.4	30.4
	otherdiscipline	25	8.0	8.0	38.5
	samearea	85	27.2	27.2	65.7
	samespecialty	58	18.6	18.6	84.3
	samespeechandlanguage	49	15.7	15.7	100.0
	Total	312	100.0	100.0	

Collaboration in the same research area (e.g. speech, or language) and to a lesser but still significant degree in the same research speciality (e.g. speech recognition or language processing) continues to be the predominant pattern at the level of the second most important collaborating research group. A similar pattern emerges for the third most important collaborating research group.

Moreover, when one analyses the specialisations of collaborating researchers in a more focused way, a similar pattern is revealed. Two separate analyses were conducted - for speech researchers and for language researchers. However, no significant statistical differences were detected between the two groups. In this regard, both communities appear to have a similar pattern of collaboration. Figure 4-2-9 shows the results for speech researchers.

Figure 4-2-9 - Specialisation Pattern of More Intensive Collaborators, at the Level of Researchers



Legend: Values in the *X* axis should be read as indicating the following categories:
Same Research Area: 1, 2, 6, 7; Same Research Speciality: 16, 17, 21, 22
Speech and Language: 19, 20, 24, 25; Other Discipline: 4, 5, 9, 10
Other values represent intermediate values between the identified groups

Here, again, the empirical evidence corroborates a pattern of relatively specialised collaboration within the same research speciality or within the same research area. The relative importance of collaboration within the research speciality increases when we analyse collaboration at the level of individual researchers.

This section concludes the analyses of patterns of research collaboration at the level of individual researchers and at the research group and institutional levels. In the next section we discuss the conceptual framework for the analysis of electronic networking of research networks, with a special focus on the identification of the *structural* components of the process of scientific collaboration and communication.

4.2.5 Electronic Networking of Researchers and Research Networks

The electronic networking of research should be intrinsically related with the structure of research collaboration. The persistent long-terms patterns of research collaboration, at its various levels of connectivity among researchers, research groups and research institutions, reveal the structure of scientific organisation (this has been empirically verified in the previous sections 4.2.3 and 4.2.4). It is the nature of this structure (in terms of specialisation and division of labour, in terms of the connectivity of the whole community or parts of it, in terms of the hierarchical structure in the organisation of the community and the differences in prestige and centrality of researchers or research groups, as well as in terms of the evolution and dynamics of collaboration links and patterns of connectivity) that determines the degree of success in exploiting the benefits of using advanced Information and Communication Technologies in science. On the other hand, the functioning of electronic collaboration and communication is also likely to bring about changes in the traditional structures of scientific collaboration and communication. The analysis of this systemic relationship is a fundamental step towards a better understanding of e-science and the patterns emerging from electronic scientific communication. In other words, it is the systemic interaction among the new technological capabilities of electronic networks - Technological Infrastructure - with the patterns of research collaboration and the processes of communication within the community - Socio-Organisational Infrastructure - that provides the optimal exploration of the benefits of ICTs for e-science.

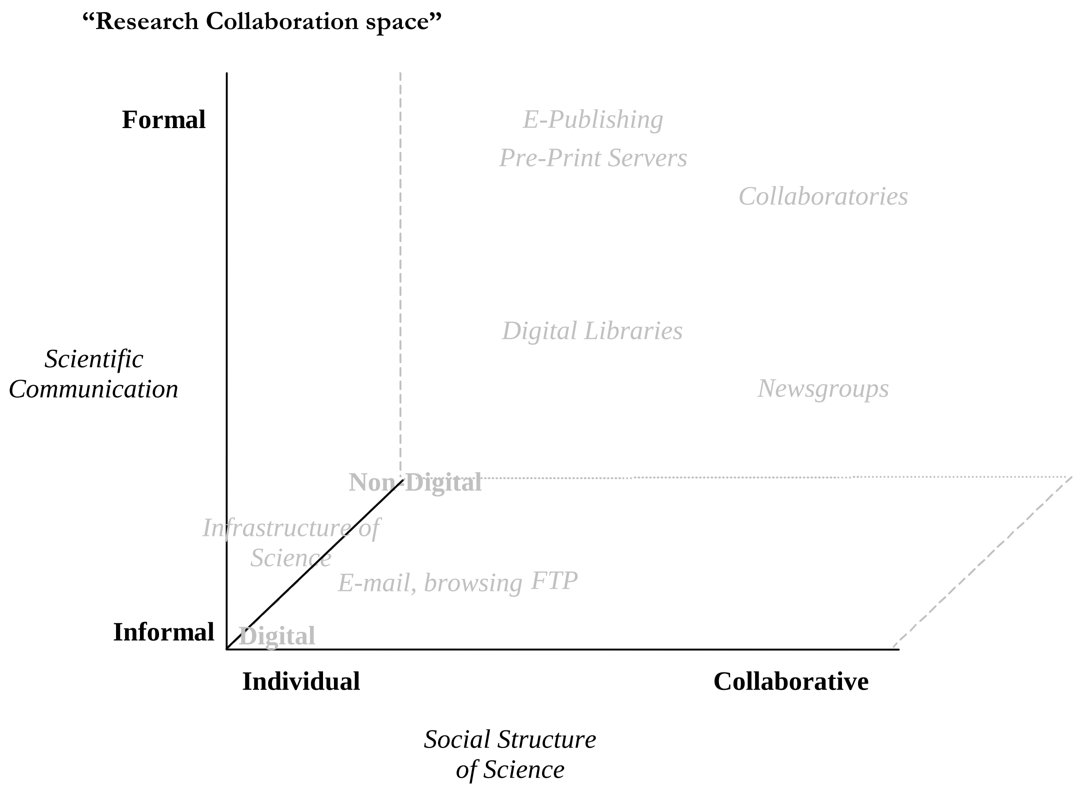
Among several other aspects of the development and implementation of digital infrastructures for scientific collaboration and communication, our principal focus is on the changing patterns of communication among scientists - particulary how and to what extent electronic networks are supporting research work, communication and collaboration. Other research investigations, not tackled here, include the provision of physical electronic networks, the development of specialised applications for scientific collaboration, providing remote access to scientific instrumentation, the exchange of scientific information through the support of digital libraries and the electronic networking of research with the education and training systems, among the more important.

When one focuses on the analysis of changing patterns of communication among researchers and other research collaboration "forms", the existence of the technological infrastructure - distributed electronic networks – is, in this investigation, taken as constituting one dimension in a wider collaboration "space".

An alternative view would have been to consider the technological infrastructure as an extended medium for the conduct and practice of research. From this alternative functionalist perspective, the traditional scientific communication model would perhaps be transformed into an "extended electronic scientific communication model". In this way, in encompassing the availability of electronic media, the research process modelled in a linear fashion from the conception of a scientific idea to the publication and dissemination of results (see the Garvey and Griffith model in the literature review - section 2.1.2), would be "extended", but in no fundamental way re-structured. Yet another perspective would have been to consider the effect of the availability of electronic infrastructures and tools focusing on the level of the individual researcher. It would be certainly necessary to analyse thoroughly how, for the individual scientist, the practices and behaviour of communication and research work change with the availability and use of these tools. However, this perspective underestimates the "collaborative" spectrum of activities, beyond the individual. Yet, a third alternative would have been a "top-down" strategy, commonly applied to science policy interventions in this regard, considering the electronic networking of research to be fundamentally focused on the provision of advanced electronic infrastructures and applications, but less so on the traditional structures of collaboration and communication within scientific communities and how these influence electronic networking.

The empirical analyses of patterns of research collaboration and communication (see sections 4.1 and 4.2, on quantitative measures of scientific communication and patterns of research collaboration and research networks, respectively) reveal a better understanding of the structures of research collaboration in speech and language. These analyses represent a less conventional approach to the study of electronic environments for scientific communication. The proposed conceptual model results not from an extension of the traditional linear scientific communication model, but from the extension of a "research collaboration space" model. Figure 4-2-10 represents schematically the essential components of such a model.

Figure 4-2-10 - "Electronic Research Collaboration" Model



Recalling the discussion in section 2.3, about the more theoretical conceptualisation of the investigation, this "electronic research collaboration" model enables the characterisation of the systemic interaction among the Socio-Organisational Infrastructure of science with the Technological Infrastructure.

First, it takes into account the multi-dimensional nature of the research collaboration process. There is one axis or dimension representing the variegated "forms" of research collaboration, at its various levels, from individual research endeavour to very complex collaborative research activity. This dimension is labelled "social structure of science", as it represents the organisation of science, and the structure of collaboration activities. There is a second complementary dimension, representing the whole spectrum of the nature of communication and collaboration activities, from the completely "informal" processes of communication to the formal communication artefacts and activities, such as the end-products of scientific communication in the form of scientific articles, or the institutionalised research network projects. As discussed earlier, particular forms of research collaboration, such as "research networks" might be represented as diverse entities along this three-dimensional spectrum - and not as a precise point located in a single dimension. Secondly, the model embeds the electronic networking of research collaboration within the structure of the process of scientific communication and collaboration. A third dimension encompassing the Technological Infrastructure of science is added to the initial "research collaboration space". Indeed, the various "forms" of electronic networking, from inter-personal communication tools such as e-mail, to highly collaborative ones, such as "newsgroups" or more extensively "collaboratories" are embedded in the research collaboration space. Here, again, the various forms of electronic networking of research are represented not as single points in space, but as volumetric "forms" across a wide spectrum.

Up to this point in the empirical analysis we have mainly focused on examining patterns of the socio-organisation of the speech and language community. The discussion to follow examines in more detail the technological infrastructure, giving particular emphasis to the three dimensions that were found to represent gaps in knowledge in terms of socio-organisational support in electronic networks: first, empirical evidence on individual use of Internet technologies and electronic networks in supporting researchers' work, communication and collaboration; secondly, electronic networks as effective infrastructures supporting research collaboration and reproducing in electronic environments traditional and new patterns of interaction; and, finally, electronic networks as infrastructures for wide dissemination and distribution of information and electronic representation of knowledge inter-change.

The first three sections of chapter 5 (sections 5.1 to 5.3) analyse empirical evidence testing the robustness and validity of this conceptual model. Empirical data are discussed focusing: 1) on the individual use by scientists of ICTs, and particularly Internet technologies, for research work, communication and collaboration; 2) analysis of electronic collaboration in interactive environments such as newsgroups; and 3) the analysis of knowledge exchange and dissemination of scientific information on the Internet, giving rise to the formation of web communities and digital knowledge bases.

4.2.6 Conclusion

In this section, the analysis has focused on the structure of research collaboration. The characterisation and detailed study of patterns of collaboration and long-term regularities in scientific communication are fundamental to an understanding of the electronic networking of research resources. We can only appreciate the changes in the digital infrastructure of science if we embed these structural changes in the patterns and structure of research collaboration and communication.

The theoretical discussion was organised in the following way. First, we looked at the understanding of research collaboration forms, such as research networks, as a form of organisation of scientific communities. Secondly, we focused on the discussion of the concept of a "research network" as one, among several other, forms of "research collaboration". Thirdly, we discussed motivations, incentives, costs and other factors explaining research collaboration practices. Moreover, we attempted to give an overview of the science policy activities endorsing research collaboration. This demonstrated how difficult might be to "measure" research collaboration, as well as the "performance" of research collaboration structures.

Within this theoretical framework, we tested hypotheses concerning patterns of research collaboration, attempting to identify long-term structures and regularities. These collaboration structures were examined at two complementary levels: 1) institutional collaboration patterns, based on empirical evidence of research funded co-operative projects in speech and language research, over the period 1990 - 2002; and 2) collaboration patterns of individual scientists, based on analysis of a survey of researchers belonging to the ELSnet network.

The empirical evidence validated some theoretical hypotheses. First, "research networks" are dynamic systems, changing continuously over time, as the evolution of various networks testify. Moreover, we need to rigorously delimit the concept of a "research network" in a two-dimensional collaboration space encompassing levels of collaboration and the formal/informal nature of the collaboration activities. Secondly, there are some long-term collaboration patterns that give rise to structures of collaboration and densely knit groups of regular collaborators. Thirdly, the position a research institution occupies in the collaboration space is relevant, particularly in terms of its centrality and connectivity. Fourthly, it does not follow that the best-connected and most central research institutions are the "centres of excellence", but potentially these are necessary conditions.

At the level of the individual researcher also, we found empirical evidence of continuing patterns of collaboration. First, we validated the international and inter-sectoral character of this community of researchers. Secondly, we examined the interdisciplinarity of the researchers' backgrounds, as well as the specialisation and division of labour of the collaboration structures. Thirdly, we tested whether collaboration patterns change according to "seniority" in research. We found that this does indeed occur. Finally, we found that personal collaboration networks may reveal important structures, above the level of the individual scientist, involving intra- and inter-institutional collaboration.

In the final section, we revised our initial conceptual framework for the analysis of changing patterns of collaboration and communication within electronic environments. This conceptual approach required the use of innovative methodologies combining both the collection of researchers' experiences in the use of Internet technologies - using traditional survey techniques, as well as the observation of electronic environments detecting structures of collaboration and communication - through development and use of innovative Cybermetric techniques. The empirical results and theoretical validation of this model are presented in chapter 5 (sections 5.1 to 5.3). But first, we attempt to synthesize in the following chart (Table 4-2-X) the potential differences and similarities among traditional (non-electronic) and electronic forms of communication and collaboration.

The table in the next page depicts some tentative similarities and differences in forms of communication in the traditional (non-electronic) and new electronic settings.
In order to facilitate the analysis of the transition to electronic environments, we present the results structured in three complementary aspects: Research Work and Group-Work; Collaboration and Social Networks; and Information Dissemination and Distribution.
When analysing differences in workgroup and research work, we are particularly emphasising properties of the internal process of communication and collaboration during the research workflow activities. New electronic networks are likely to allow remote workgroup communication and access to advanced tools and resources in distributed locations. The analysis of research collaboration structures and social networks emphasises the organisational forms resulting from communication and collaboration (whether in the traditional – non-electronic settings – whether in the new electronic networks). Finally, when analysing similarities and differences in forms of information dissemination we are particularly interested in how research information assumes different formats and supports different functionalities. Empirical confirmation is still missing for the electronic settings.

Table 4-2-X – Comparison of Traditional to Electronic forms of Communication and Collaboration – Initial proposal

Traditional Research Communication and Collaboration	Electronic Research Communication and Collaboration
Research Communication and Group-Work	
- Highly localised and bounded workgroups	- Highly distributed but limited e-mail Bulletin Board's workgroups
- Geographically localised workgroups	- Geographically remote and distributed workgroups
- Specialised Workgroups	- Electronically specialised workgroups
- Linear research process workgroup (from initial idea to project's results)	- "Complex systems" workgroup research GRID architectures, resources and tools
Collaboration Structures and Social Networks	
- "Invisible Colleges" – close knit and tightly connected groups of peers	- Electronic Invisible Colleges – Newsgroup invisible colleges or E-Mail Peer's communities; Pre-Print Server's Communities
- Research Groups and Department Groups	- Remote "extended research groups" and "mission-specific" electronic research groups
- Informal Social Networks	- Newsgroups, electronic links among researchers and research groups, electronic "webs of assistance"
- Institutionalised Research Networks	- Remote Servers with network specific resources; Collaboratories
Information Dissemination and Distribution	
- Preliminary ideas and research proposals	- Personal e-mail documentation
- Meeting reports and documentation	- Mailing lists and FTP or Remote Servers
- Conference and Symposia reports, articles and proceedings	- Internet Conference tools and repositories of information, Digital Libraries
- Research projects documentation	- Electronic research projects (with modularity*)
- Personnel and institutional information	- Personal and Institutional web presence
- Pre-print publications	- E-Print Archives (with modularity*)
- Journal articles	- Electronic Journals and E-Publications (with modularity*)
- Books	- Electronic Books (with modularity*)

* Modularity allows the computation of parts and components of documents in various formats.

The following chapter discusses results of the empirical evidence analysed in the electronic communication and collaboration of speech and language researchers and research groups, in order to validade the suggested patterns of change.

5. Data Analysis and Results - Electronic Scientific Communication and Collaboration

In this chapter we analyse the empirical evidence on patterns of electronic scientific communication in computational speech and language and the interactions of the new technological infrastructures resulting from the availability of distributed electronic networks with the socio-organisation of researchers and research groups in this research community.

The particular interest is examination of the empirical data to address the research question of whether or not - and to what extent - traditional (non-electronic) features of scientific communication and patterns of research collaboration (examined in previous sections) are being reproduced in these electronic networks. Having identified and characterised the socio-organisation of the speech and language community, and particularly researchers and research groups within the European area, in this chapter we analyse the characteristics of the technological infrastructure and its adoption for networking, communication and collaboration.

The conceptual framework (discussed in chapter 2) identified a major gap in the theoretical and empirical knowledge within the general theme of e-science and digital infrastructures for science: for example, how are electronic networks being used in terms of supporting social networks (in our case science networks)? This investigation aims explicitly to contribute to advancing the knowledge in this regard. Moreover, within the analysis of this systemic techno-social interation we focus on three complementary issues: individual use by researchers of Internet technologies for research work, communication and collaboration; the extent to which electronic environments support the reproduction of traditional patterns of collaboration and features of formal and informal communication and interactions; and the extent to which these electronic networks facilitate the distribution and disseminaton of scientific information and the construction of large-scale and distributed knowledge networks. In particular, we emphasise the socio-organisational adoption of these technological infrastructures.

Section 5.1 discusses in detail, using evidence from the survey, how researchers use these Internet technologies and ICTs in general for conducting research work, and particularly for communication and collaboration. We seek a better understanding of the how much these electronic networks support formal and informal processes of communication and at the same time effectively support collaborative interactions - whether small-scale inter-personal collaboration or highly collaborative and distributed activities. Empirical evidence reveals the extent of multi-modality required of these technologies for different dimensions of research activity.

In section 5.2 we examine how electronic networks - and particularly technologies suitable for distributed and interactive communication, such as Newsgroups - support informal scientific communication, reproduce features of communication within scientific communities and extend patterns of non-electronic socio-organisation (hierarchical structures, "invisible colleges", specialisation and division of labour, inter-sectoral and international collaboration) into electronic environments. Moreover, we attempt to identify the kind of information and knowledge which is being exchanged within these more interactive electronic networks.

Finally, section 5.3 focuses on analysis of the extent to which Internet technologies support the dissemination of information and knowledge in a non-randomised or chaotic system, but also reproduce communicative and collaborative interaction extended electronically from non-electronic forms of organisation. We analyse patterns of electronic connectivity as revealed by the inter-linkages of information distributed on the Internet and test the hypothesis of its similarity with traditional (non-electronic) patterns of collaboration among the research groups. In a more exploratory way, we identify a procedure to characterise electronic or digital knowledge bases – that is, large scale and distributed electronic knowledge structures.

The above discussions throw light on our initial conceptual framework, providing empirical evidence on the extent to which electronic networks reproduce traditional (non-electronic) patterns of communication and collaboration (see final synthesis in the conclusions, chapter 6).

5.1 Internet Technologies and the Process of Scientific Communication and Collaboration

5.1.1 Introduction

In this section, we analyse the direct influence of the use of ICTs on the process of scientific communication and collaboration. We will be particularly focusing on Internet usage patterns within a particular scientific community. Our primary aim is to assess the influence of the use of these technologies on patterns of communication among scientists. The empirical data result from an electronic survey of the European Speech and Language research network - ELSnet (see chapter 3 about methodology).

The overall effect of the use of ICT on scientific communication is strongly determined by or even dependent on structural patterns of collaboration within scientific communities as well as on the peculiar characteristics of the formation and development of research fields. In section 4.1, we analysed important stages in the formation and development of the speech and language research area using bibliometric analysis of publications. Moreover, the social context characterising the organisation of academic communities is a key determinant of the adoption of particular technologies as a medium of enhanced communication. In section 4.2, we analysed important characteristics of the collaboration process in this particular field of research. Science policy initiatives have had a significant influence on the interdisciplinary nature of collaboration and its international and inter-sectoral patterns, at the level of the individual researcher, as well as the institutional levels of research institutions and whole research networks.

The adoption of ICT for scientific communication and collaboration is also partly dependent on the technological characteristics of these communication media and their development over time. From the diverse set of ICTs that is available, Internet technologies are particularly suited for our analysis, as they are built upon networking services and protocols, are extensively based on open standards and widely support open distribution of information. This technological environment is fundamental for the networking of resources, information and knowledge (see section 5.3 on the discovery of Digital Knowledge Bases) and enables the networking of people and organisations (researchers and research institutions). These electronic networks are indeed natural environments for collaboration. These collaborations assume the form of informal and interactive collaboration (see section 5.2 on Newsgroups), as well as more formal architectures (see section 5.1.7 below on specialised infrastructures for scientific communication).

Some important operational hypotheses about this interplay of ICT and patterns of collaboration and scientific communication are worth empirically testing. First, we would expect these technological environments, compared to traditional media of communication (fax, phone and face-to-face), to be particularly suited to remote collaboration and remote interdisciplinary communication within an interdisciplinary research area within the speciality. In addition, the importance of these technologies for local collaboration should be empirically tested.

Secondly, the intensity of use of different Internet services (e-mail, Web, FTP, Print-Servers, etc.) is likely to vary between researchers. We explore the hypothesis that this intensity might somehow be correlated with the "seniority" of researchers, as given by number of years in research, or their "scientific productivity" as given by number of publications. Moreover, we introduce the concept of "multi-modality" as the norm in the use of different technological services for different stages of the research process.

Thirdly, as these electronic networks allow extensive sharing of resources, information and knowledge, we explore whether or not different "scientific information", such as staff contacts, research projects, research publications and links to other closely related research groups, is widely distributed or disseminated within these networks. The data are discussed at the level of the individual researcher as well as at the level of the research group.

Finally, we test whether researchers use in an intensive way specialised electronic infrastructures and services, such as access to remote computational facilities and the direct use of pre-print servers.

The analysis of patterns of ICT use and the empirical evidence offered by researchers participating in the ELSnet network are organised in the following way. In section 5.1.2, we discuss electronic communication as a medium for "Extended Research Networks", allowing remote and interdisciplinary collaboration as well as intensive local collaboration. Section 5.1.3 discusses the "multi-modality" of different technologies for different stages of the research process and patterns and intensity of ICT use. Electronic networks as repositories of knowledge and environments for knowledge-interchange are discussed in section 5.1.4. Finally, in section 5.1.5 the discussion focuses on specialised and advanced infrastructures for scientific communication and collaboration, such as access to high-performance and distributed computing resources, electronic scientific publishing, “collaboratories” and GRID architectures.

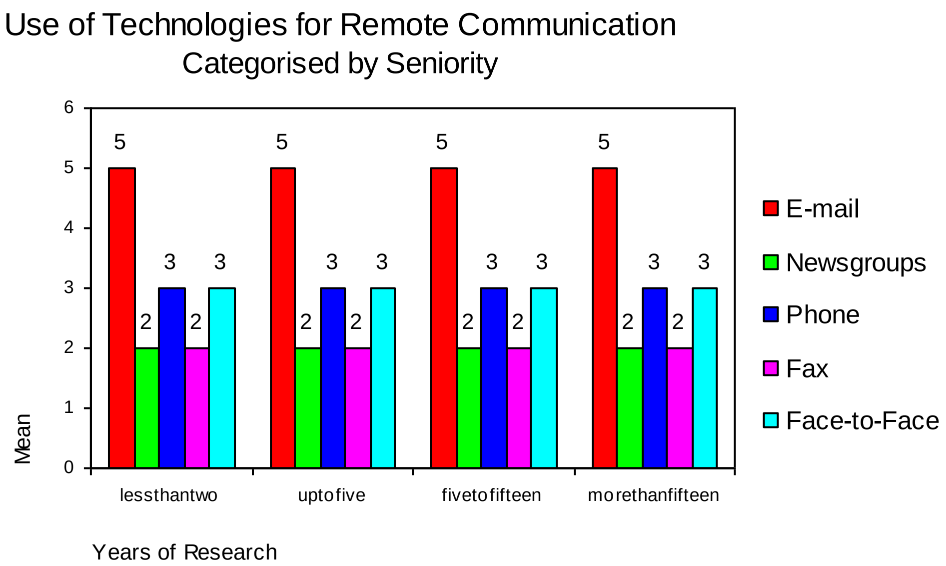
5.1.2 Electronic communication as a medium for “Extended Research Networks”

In this section we analyse, from the viewpoint of researchers, the relative importance given to these Internet inter-personal technologies for local/remote collaboration, as well as for collaboration within one speciality of research and with other disciplines compared to more traditional means of communication (phone, fax and face-to-face). Theoretically we would expect these technologies to contribute more to “extended research groups” – groups collaborating at a distance, but at the same time being well connected and effective.

One result from the survey is that the relative importance of using these ICTs for purposes of collaboration is not associated with the “seniority” of the researchers. This result is valid for the various dimensions of the analysis (local or remote collaboration; collaboration with close colleagues or other colleagues; and collaboration within the speciality of research or with other disciplines). Another general result from the survey is that some traditional technologies (e.g. fax) are considered insignificant for the purpose of communication in any of the above dimensions.

The empirical data corroborate the importance of these technologies (particularly e-mail) for remote collaboration. In fact, a major proportion of the respondents (93%) rated the use of E-mail as “very important” for remote collaboration irrespective of whether for communicating with close/other colleagues, within the speciality or from other disciplines. Figure 5-1-1 depicts the relative importance of new/traditional technologies for the purpose of remote collaboration.

Figure 5-1-1 - Relative Importance of Internet Technologies for Remote Collaboration

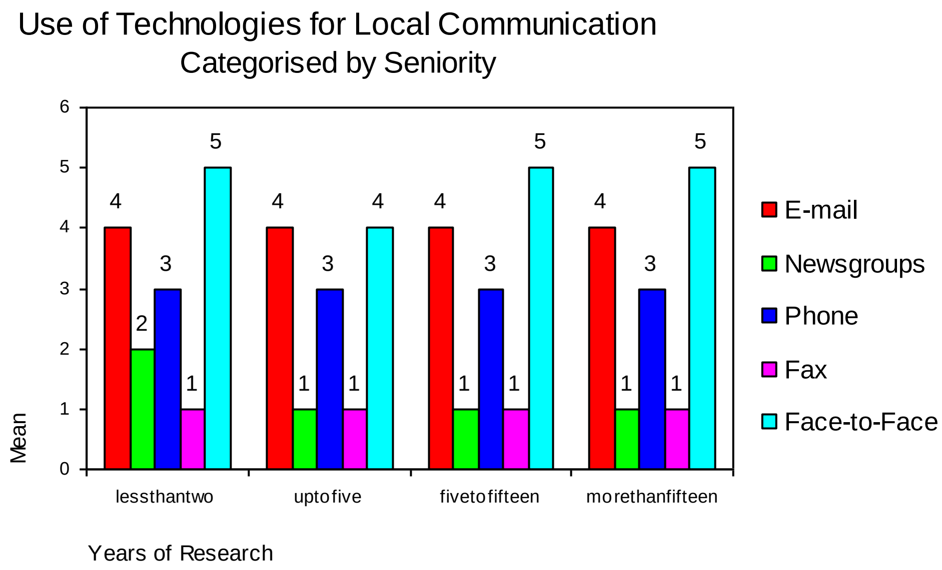


Legend:
5 - Very Important 4 - Important 3 - Average 2 - Relatively Insignificant 1 - Non-Significant

It is worth noting that, while E-mail ranks very highly as an inter-personal remote collaboration tool, the same is not true for Newsgroups communication. Newsgroups are rated as relatively insignificant for collaboration in the various dimensions, except for collaboration within the speciality for some groups. This is probably related to the public nature of these electronic systems and the fact that they are in fact not inter-personal collaboration tools but collective collaboration tools (see section 5.2 for extensive discussion of Newsgroup technology).

On the other hand, the data also reveal the importance of traditional face-to-face communication, particularly for local communication and communication with close colleagues. For local communication, all but one group – that of up-to-five years' research experience - ranked it as very important. And when considering collaboration with close colleagues, all groups of researchers without exception ranked face-to-face communication as very important. These results are depicted in Figure 5-1-2. For both dimensions, face-to-face is relatively more important than E-mail technology.

Figure 5-1-2 - Face-to-face Communication is Still Determinant for Local and "Collegiate" Collaboration



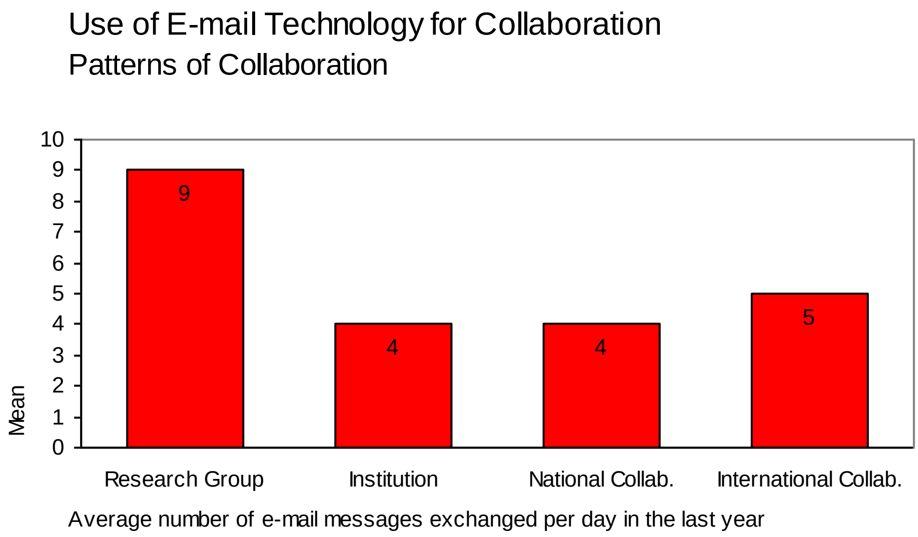
Legend:
5 - Very Important 4 - Important 3 - Average 2 - Relatively Insignificant 1 - Non-Significant

When one analyses the intensity of use of E-mail technology for collaborative activities with various groups of researchers (researchers from the same research group, from the same institution, other national researchers, or other international researchers), some interesting results emerged. While the importance of this technology for remote collaboration was reinforced, its importance for local collaboration within the research group was also stressed.

This result is of special relevance for the linkage between local research group collaboration and remote collaboration with the international community. It is not a sufficient but is an essential and fundamental condition for "extended research groups" via electronic networks.

Figure 5-1-3 represents the above results graphically.

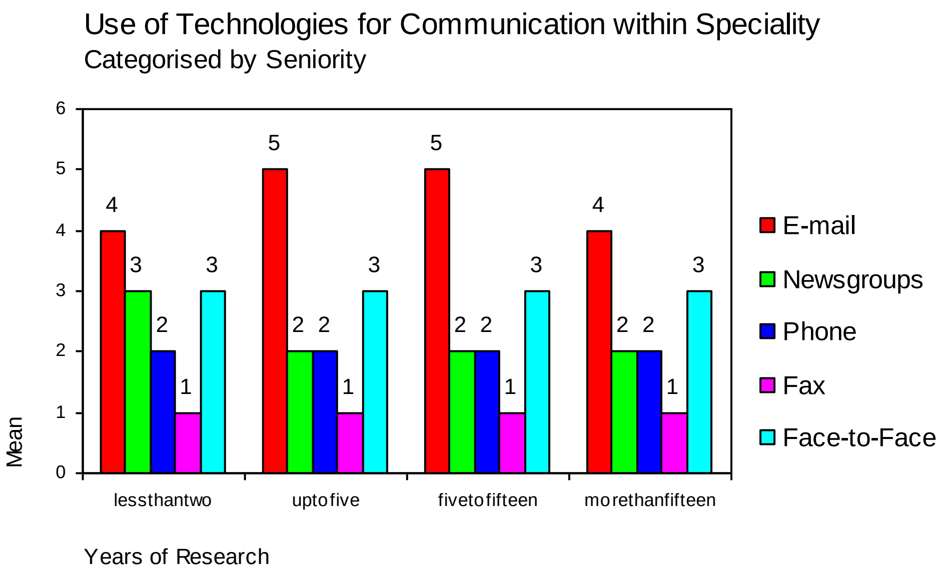
Figure 5-1-3 - Importance of E-mail Technology for Research Group Collaboration



A final dimension that was analysed was whether or not the pattern of relative importance attributed to these electronic technologies varies in a significant way for collaboration within the same field of research or with researchers from other disciplines.

Figure 5-1-4 presents the results for the importance of these technologies within the same field of research.

Figure 5-1-4 - Relative Importance of Communication Technologies within the same Research field



Legend: 5 - Very Important 4 - Important 3 - Average 2 - Relatively Insignificant 1 - Non-Significant

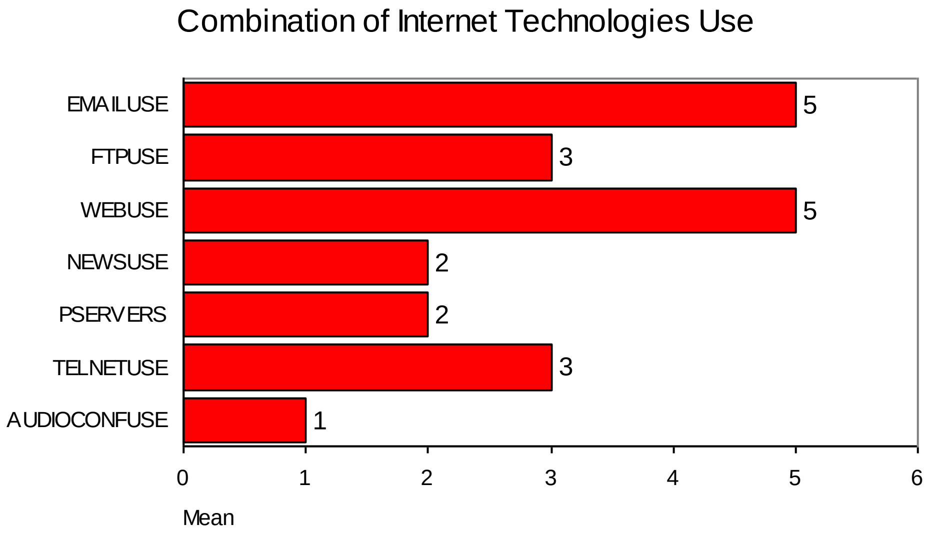
No statistical difference was found in the relative importance given to any of these technologies considered individually and as a group, for purposes of collaboration within the same research area or with other scientific communities. Moreover, no significant correlation was found between the degree of "specialisation" of collaboration (as given by the researchers network of close researchers disciplinary affiliation) and the importance of communication technologies for collaboration.

5.1.3 Multi-modality in the Use of Different Technologies

In this section we will focus on the intensity of use of different electronic communication technologies for several stages of the research process. Moreover, we will determine whether or not the pattern or combination of using of these technologies is related to "seniority" in research or to "research productivity". First, we will examine how intensively and how differently these ICT technologies are being used for research work. Secondly, we will test the hypotheses of correlation between the intensity of use of these technologies and "seniority" and "research productivity". Then, the "learning process" involved in using these technologies is examined and finally we focus on the use of different technologies for different aspects of the research process.

The intensity and frequency with which researchers use these different Internet technologies varies significantly. While E-mail and Web technology are widely and intensively used by the great majority of the respondents (87% reported using E-mail and 83% using the Web, daily and intensively), other technologies - particularly Audio-conferencing - are rarely (20%) or never (60%) used. Two other categories of technologies in terms of frequency and intensity of use were empirically revealed. File Transfer Protocol (FTP) and Remote Servers (Telnet), and Newsgroups and Print Servers. The former category is used more often and in a more intensive way, while the latter category - Newsgroups and Print Servers – were categorised as being not used intensively by a significant proportion of researchers. In fact, 60% of respondents reported using, with some frequency and intensity, FTP and 41% Telnet technology while Newsgroups are never used or rarely used by 58% of respondents and Print Servers by never or only rarely used by 53%. Figure 5-1-5 summarises these results.

Figure 5-1-5 - Intensity and Frequency of Use of Different Internet Technologies



Legend:
Free Scale 5 - Daily and Intensively .... 1 - Never

The above results highlight the importance of distinguishing among the different technologies, given the significant variance in the patterns of their use for research work.

A more interesting theoretical hypothesis concerns the potential association between different levels of "seniority" in research (seniority as measured by the number of years in research) and the intensity of use of these modern electronic technologies. No significant correlation was detected between intensity of use of these technologies and "seniority". The analysis was pursued for each type of technology - from E-mail to Audio-Conferencing, and each "seniority" group was cross tabulated against each different level of intensity of technology use. The results were quite consistent and robust over all technologies and seniority groups. Table 5.1.I presents the results for use of Web technology.

Table 5-1-I - Cross Tabulation Comparing Use of Web Technology with "Seniority" of Research, as Given by Number of Years in the Research Area.

Crosstab

			yearsresearch				
			fivetofifteen	lessthantwo	morethan fifteen	uptofive	Total
WEBUSE	1.00	Count			1		1
		% within WEBUSE			100.0%		100.0%
		% within yearsresearch			1.6%		.3%
		% of Total			.3%		.3%
	2.00	Count	3	1	1	1	6
		% within WEBUSE	50.0%	16.7%	16.7%	16.7%	100.0%
		% within yearsresearch	2.7%	2.4%	1.6%	1.2%	2.0%
		% of Total	1.0%	.3%	.3%	.3%	2.0%
	3.00	Count	9	7	12	7	35
		% within WEBUSE	25.7%	20.0%	34.3%	20.0%	100.0%
		% within yearsresearch	8.1%	17.1%	18.8%	8.3%	11.7%
		% of Total	3.0%	2.3%	4.0%	2.3%	11.7%
	4.00	Count	20	9	15	24	68
		% within WEBUSE	29.4%	13.2%	22.1%	35.3%	100.0%
		% within yearsresearch	18.0%	22.0%	23.4%	28.6%	22.7%
		% of Total	6.7%	3.0%	5.0%	8.0%	22.7%
	5.00	Count	79	24	35	52	190
		% within WEBUSE	41.6%	12.6%	18.4%	27.4%	100.0%
		% within yearsresearch	71.2%	58.5%	54.7%	61.9%	63.3%
		% of Total	26.3%	8.0%	11.7%	17.3%	63.3%
Total		Count	111	41	64	84	300
		% within WEBUSE	37.0%	13.7%	21.3%	28.0%	100.0%
		% within yearsresearch	100.0%	100.0%	100.0%	100.0%	100.0%
		% of Total	37.0%	13.7%	21.3%	28.0%	100.0%

SPSS result for crosstab of 300 respondents, Chi-Square Sign. = 0.270

A similar statistical analysis was conducted to compare intensity of use of these different technologies with "research productivity" of researchers (as given by two "proxy" measures - total number of journal articles published and number of papers presented in conferences). Again, no significant association was detected between intensity of use of these technologies at any level and "research productivity".

Nevertheless, the empirical results revealed a "learning process" in the adoption of these technologies in terms of an evolutionary pattern in their adoption, from the technologies used less intensively and frequently to those used regularly in day-to-day research work.

Figure 5-1-6 - Time of Adoption of Print Server Technology

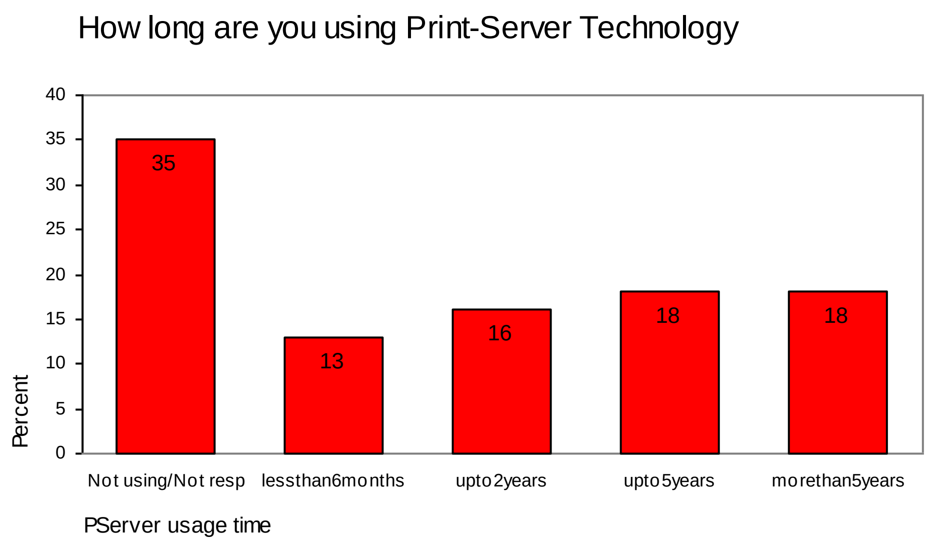
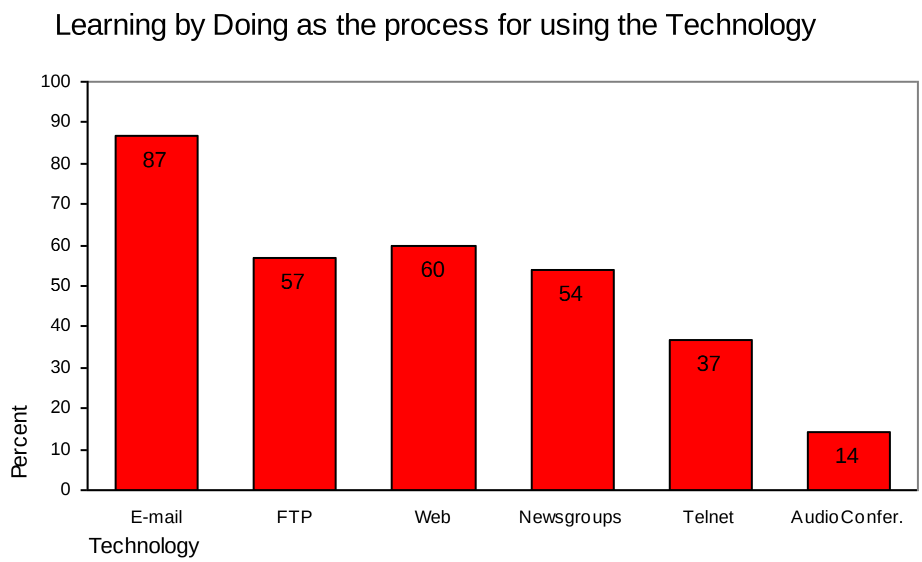


Figure 5-1-6 above shows that at least 13% of the respondents have been using this PrintServer technology for less than six months and in addition 16% learned how to use it less than two years ago. Newsgroup technology, on the other hand, has been used for a long time. Only 20% of respondents learned to use this technology in the past two years. About 38% have been using the technology for more than five years and 21% have been using it from between two and five years. However, technologies currently being used more intensively and more frequently, have also been used for a reasonably long period of time now. For example, 71% of respondents have been using E-mail for more than five years, and 17% have been using it for between two and five years. Following a similar pattern, 55% have used Web technology for more than five years and at least 31% have been using it for between two and five years. For FTP, 59% of respondents reported that they had used it for more than five years and 20% for up to five years. This implicit "learning process" might mean that in the future there is likely to be more intensive use of technologies such as Newsgroups and Print-server technology. Figure 5-1-7 depicts these results.

Figure 5-1-7 - Learning by Doing in Using Internet Technologies



Another aspect of the "learning process" is the mechanism by which researchers learn to use these technologies. The empirical data confirm the hypothesis that learning–by-doing is the most common way of learning to use these electronic tools. In fact, learning-by-doing is reported as being the process most used to learn about these technologies, over other alternative processes, such as learning from colleagues, learning through professional training, and using on-line or off-line tutorials. Eleven percent of the population did not respond to this particular question. For example, 87% of respondents reported learning to use E-mail technology by doing, 57% for FTP and 60% for Newsgroups. Guidance from a colleague also figured in learning how to use RemoteServers (16%) and FTP technology (23%).

The next topic is related to what these technologies are being used for and what is their relevance for different aspects of research communication. Table 5-1-II summarises the relative importance of different technologies for various aspects of the research communication process.

Table 5-1-II - Comparative Analysis of the Importance of Internet Technologies for Research Work

Importance of different Internet Technologies for Research Work	Percentage of users who mark these technologies as important					
	E-mail	FTPServers	Web	Newsgroups	PrintServers	RemoteServers
Co-Authoring publications	88%	24%	19%	4%	5%	9%
Working in Projects	85%	25%	34%	3%	5%	12%
Organising Seminars, Conferences, ...	78%	15%	35%	11%	2%	8%
Accessing on-line Journals	13%	18%	79%	7%	15%	15%
Exchanging results with close collaborators	81%	32%	24%	3%	3%	11%
Exchanging results with scientific community	49%	29%	51%	16%	9%	9%
Disclosing working research	42%	18%	55%	12%	10%	6%
Accessing other complementary specialities work	17%	13%	81%	13%	9%	9%
Accessing other disciplines work	16%	11%	80%	12%	8%	8%
Accessing information on Funding	29%	7%	74%	10%	2%	4%
Accessing new contacts	42%	5%	67%	13%	2%	4%

Different technologies are being used in specific ways for different aspects of research work. The recognition of this *multi-modality* in the use of Internet Technology for research collaboration and communication is important when assessing the overall influence of ICT technology on the science system.

Technologies more suited to personal inter-change of information (e.g. E-mail), while being widely regarded as relevant for most activities involved in the research process, are relatively more important for "active" processes, such as co-authoring, research work on projects, exchange of results with close collaborators and the organisation of seminars and conferences, but less relevant in other more "passive" activities such as accessing information about funding, research work by others or on-line publications. Other technologies regarded as important for various research work tasks, but more passive from the viewpoint of individuals (e.g. the Web) are considered as basic electronic information resources but not important for personal interchanges. It is also worth noting that access to electronic publishing (in on-line journals) is already regarded as very significant through the use of Web technology and other less user-friendly technologies, such as print-servers, remote servers and FTP servers. These latter technologies are generally considered as being useful only for online publishing rather than research. The importance of FTP technology and partially remote servers is their use for the exchange of research results with close collaborators and the scientific community at large, giving empirical support to the idea that, even if research results are not available in open friendly environments such as the Web, access to these resources can be achieved through other technological media. This applies to access to the results of work in progress in the case of Print Servers.

Even though they are considered to be an inter-personal technology, although of a more collective nature, Newsgroups are considered relatively useful for exchanging research work within the scientific community, disclosing one's research results and accessing the work of other disciplines. The empirical evidence is consistent with the more extensive analysis of these especially interactive electronic systems, which will be discussed later (see section 5.2).

From the preceding analyses the most important findings can be summarised as follows. First, among the group of "Internet Technologies", frequency and intensity of use of different electronic technologies are likely to vary significantly. E-mail and Web technology are used very frequently and intensively while audio-conferencing is almost never used. FTP and Telnet technology are more regularly used than Newsgroups, Print Servers or Remote Servers.

Secondly, we have to reject the hypotheses that neither individual use of any of these technologies nor the pattern of their combined use, is in any way related to "seniority" in research or to "scientific productivity". The empirical evidence consistently refutes these hypotheses.

Thirdly, the empirical data reveal the importance of "learning-by-doing" as the main way of learning how to use most of these technologies. Moreover, there is a "learning process" in the adoption of these electronic services, with a significant association between current intensity of usage and timing of its adoption.

Finally, and probably most importantly, there appears to be a significant *multi-modality* of different technologies for different activities involved in the research process and scientific communication. Technologies are used differently for different aspects of the research process. The relative importance of each technology for different characteristics and needs of specific research activities is corroborated. This "matrix" of technological applicability to different aspects of research work provides the empirical basis for a conceptual model combining complementary axes: the individual - collaborative nature of the research activity, the formal or informal nature of the information being exchanged and the digital nature of the infrastructure for collaboration and communication (an extensive discussion of this model is provided in section 5.4).

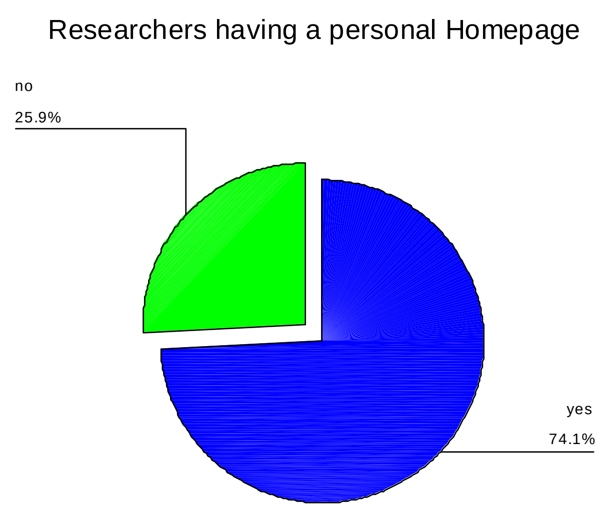
In section 5.1.5, we will explore more extensively one particular feature of these electronic environments (namely the Web), and how it is becoming a huge repository of research information, whether at the individual level of the researcher, or at the institutional level of research groups and organisations or at the level of whole research networks. Section 5.1.6 will focus on analysing specific infrastructures for electronic scientific collaboration, such as advanced and distributed computing resources and the more recent national and international research networks, "collaboratories" and GRID architectures.

5.1.4 Electronic Networks as Repositories of Knowledge and Environments for Knowledge-Interchange

This section focuses on the analysis of the wealth of research information available on these electronic networks, particularly the World Wide Web, and how it might lead to the accumulation of widely available knowledge resources in digital format as well as perhaps to creative explorations and knowledge-interchange.

The discussion is based on empirical data given by researchers about their "own" presence in these electronic environments - possession of an individual homepage and the timing of its adoption; existence of a research group/departmental presence on the Web and date of its conception; and the nature of the content of the resources disseminated on these electronic networks. Researchers were asked whether they had a personal homepage, publicly available on the Internet. The results are shown in Figure 5-1-8 below.

Figure 5-1-8 - Individual Researcher's Connectivity (Availability of Personal Homepage)

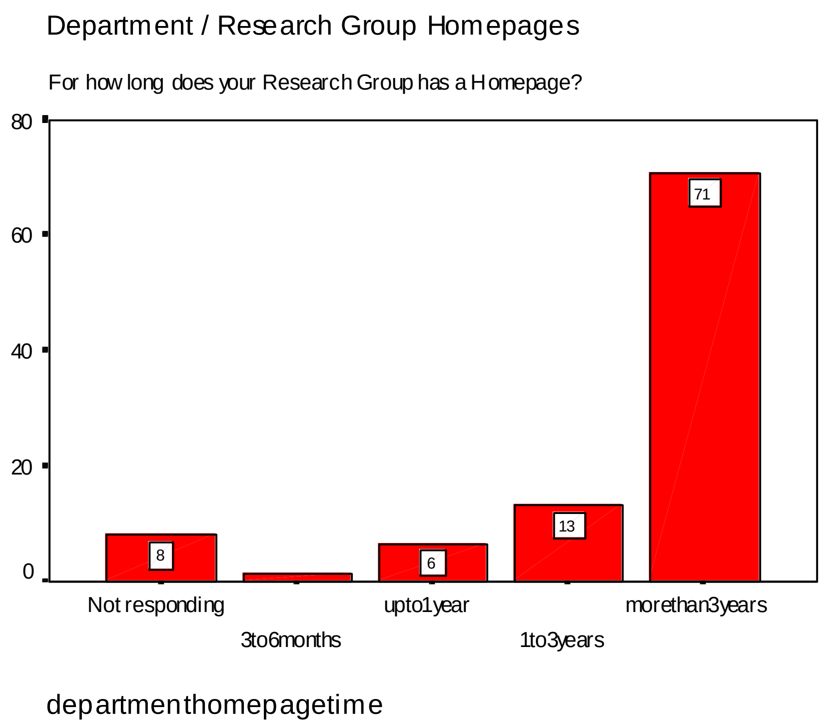


The great majority of researchers (almost 75%) currently have a presence on this electronic network and make personal information about current research and contact information available. A significant proportion (69%) have made this information available for more than three years, and 13% have had a home page including this information for between one and three years. From these data we may conclude that a significant amount of research information, at the individual level of the researcher, is currently available in these world-wide electronic networks.

When analysing the public dissemination of research information, at the institutional level of research group or department, the results were similar. Based on reporting by researchers, currently 80% of the research groups have an institutional homepage. A more detailed examination of the contents of these homepages is conducted later in this section.

Similar to individual homepages, institutional research group homepages have been disseminated on these electronic networks for some time. Figure 5.1.9 depicts time of availability of institutional homepages.

Figure 5-1-9 - Availability of Institutional Information on the Web



For 71% of the respondents, these resources have been available electronically for more than three years, and for a further 13% they have been available for between one and three years.

Recognition of the availability of these research resources on digital web space led to a more extensive analysis of the nature of the content being disseminated in these electronic networks. Table 5-1-III summarises the results concerning the type of information available as well as the frequency of update of content.

Table 5-1-III - Research Group Homepages, their Content Nature and Frequency of Update

Infromation and Contents of Research Groups Homepages - Update Frequency	Once a Month	Every 3 Months	Once a Year	Non-existent	Non-response
Institutional Contacts and Information	55	67	127	9	54
Research Group Contacts	34	85	115	18	60
Personal Homepages	46	94	86	24	62
Research Projects	38	106	100	10	58
Scientific Publications	72	92	76	15	57
Electronically distributed completed papers	62	84	60	37	69
Lectures and Educational Information	60	72	62	46	72
Links to specialised on-line Journals	17	42	76	102	75
Links to Pre-Prints, remote servers, other docs.	22	44	80	79	87
Links to closely connected Research Groups	14	60	125	41	72
Links to other Research Groups and Resources	12	56	125	37	82

Table computes percentage of respondents that reported availability and frequency of update for each content category.

Information about institutional and research group contacts, as well as information on research projects and scientific publications, are the kind of resources considered to be extensively available on these institutional homepages (less than 7% of respondents reported that they were not available from their research group electronic presence). However, and quite unfortunately, links to other more specialised "knowledge intensive" resources, such as electronically distributed complete papers, information on lectures and other educational resources, links to specialised on-line journals, or links to pre-print and remote servers were non-existent for 15%, 19%, 43% and 35% respectively of respondents.

Other significant results are worth pointing out. Information on research projects is not only extensively available electronically, but is frequently updated. Respondents reported this kind of information as being updated at least once a year (39%), and some of them (42%) reported the update frequency to be three monthly. In a similar way, information on scientific publications and electronic papers is also frequently updated. A reasonable number of respondents (28%) reported updating of information on scientific publications to be monthly. For electronically distributed papers this figure is similar (26%). Moreover 36% and 35% respectively of respondents consider that Web page contents are updated at least every three months. A final comment relates to links to closely collaborating research groups and other research groups and resources, which 52% and 54% respectively reported as being updated at least every year.

As a general point, we should consider that the institutional presence of research groups on the Web allows the dissemination of considerable information directly linked to the research activities conducted by those groups. Moreover, as this content is previously selected and filtered by individual research organisations, we would expect there to be a self-controlling mechanism ensuring "quality".

Given the empirical evidence discussed above, the combined analysis of the electronic availability of individual researchers' information, on the one hand, and institutional resources, on the other, allows these public electronic networks (particularly the Web) to be considered to be potential environments for the distribution of knowledge resources. Nevertheless, this hypothesis needs to be verified by a rigorous content analysis of specific zones of these electronic networks, something which is outside the scope of this present research. In section 5.3, we provide a more extensive discussion of these electronic environments for supporting knowledge distribution and explore further the links between research resources and institutions.

In the next section we will focus on the use of particular infrastructures for scientific networking, such as more advanced remote computer facilities. Their specificity emerges from their being particularly suited for special-purpose tasks of research work, such as high-performance and distributed computing.

5.1.5 Advanced Infrastructures for Electronic Scientific Communication

The particularity of the computational tools and electronic networks discussed in this section is that they are specifically adapted for scientific work, scientific communication and collaboration. The survey data allow one to analyse the use by researchers of advanced computing platforms for speech and language processing and their accessing of large remote databases, as well as access to dedicated computer servers for electronic sharing of publications (print-servers, remote servers), and also more advanced computer platforms for collaboration and networking. The results of the use of such electronic systems are discussed in the following paragraphs.

From the total sample of 312, 138 respondents stated that they had access to remote computational facilities. The frequency of use of those different resources is shown in Table 5-1-IV.

Table 5-1-IV - Use of Remote Computational Facilities

Accessing and Using Specific Computing Resources for Scientific Work and Collaboration	Weekly	Monthly	Quarterly	Annually	No-Use	No-response
Remote Computing for Heavy S&L processing	27%	13%	8%	7%	38%	8%
Remote Databases of S&L	12%	18%	11%	7%	42%	10%
Remote Servers to Download S&L tools	11%	12%	25%	16%	27%	9%
OnLine Documentaiton Sources	28%	28%	17%	5%	14%	7%
Remote Servers supporting Collaboration Networks	14%	9%	12%	8%	44%	14%

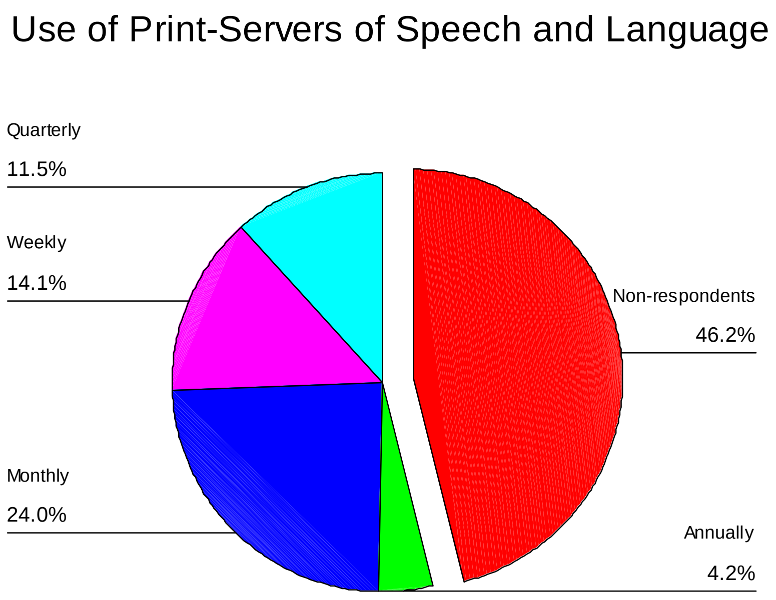
At least 27% of the respondents regularly (weekly) use high-performance computing platforms for speech and language processing. Another 13% use these computer systems once a month. Remote databases of Speech and Language information are also accessed frequently (12% use them weekly and another 18% monthly). In addition, 25% of respondents accessed remote servers to download Speech and Language special-purpose tools every three months. These results reveal that some tasks involved in speech and language research activity not only require high-performance computing, but also access to large databases of language resources, distributed over geographically dispersed sources.

Access to specific on-line documentation is also frequently required. A total of 73% of respondents access these electronic documentation sources on a regular basis (28% weekly, 28% monthly and 17% quarterly).

This information can be supplemented by analysis of researchers' access to electronic scientific publications (pre-print and post-print server resources, electronic journals).

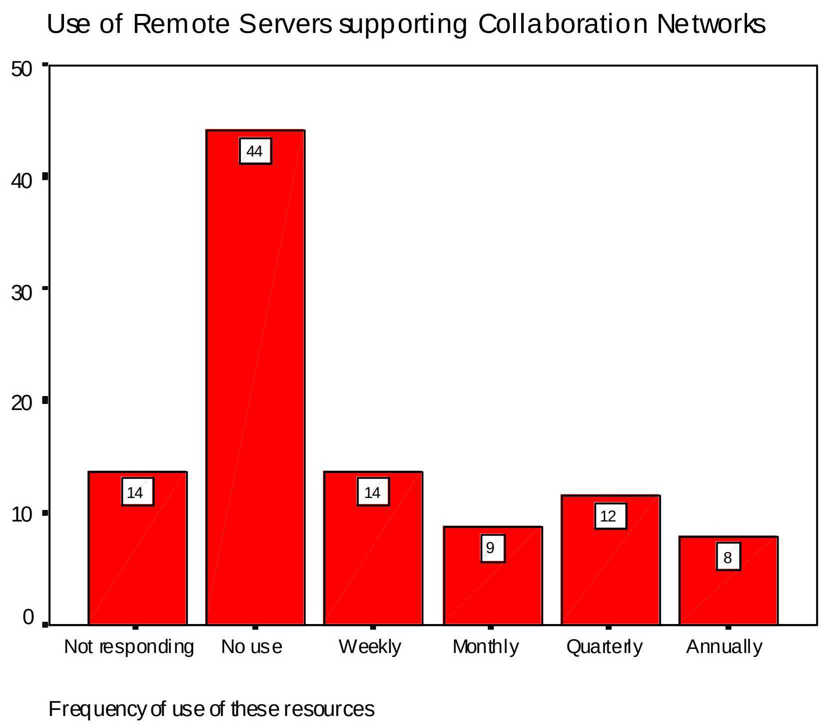
Even taking into account that 46% of the total sample did not respond to this specific question or do not use pre-print servers or remote servers for on-line documentation on Speech and Language, 38% of respondents report accessing these technologies at least monthly. This is an important indicator of the specific needs of "knowledge" resources that have to be accessed remotely. Figure 5-1-10 plots the results.

Figure 5-1-10 - Usage Patterns of Pre-print Servers of Speech and Language



When the needs and effective use of simultaneously more advanced computing resources and electronic environments specifically supporting collaboration networks are analysed, the results indicate the specialised needs of speech and language research activity. Figure 5-1-11 shows these results.

Figure 5-1-11 - Usage Patterns of Remote Servers Supporting Collaboration Networks



Although 44% of the 312 respondents reported not using these technologies and an additional 14% did not respond to this question, nevertheless, a reasonable percentage of 14%, 9% and 12% of the researchers reported using the collaboration technologies very frequently (weekly, monthly and quarterly respectively). Again, this is indicative of the special research activities applicable to the speech and language community (and as such probably not typical of other research communities).

The more integrated use of some of the above technologies as well as other electronic environments has recently received some attention. The specific needs of research communities in terms of special dedicated computational resources and restricted electronic networks, have been subject to intensive science policy discussion followed by effective action. Three of the most visible of these science policy efforts have been focused on the implementation of "National Research and Education Networks", the conceptualisation and implementation of "Collaboratories" and the technical implementation of "GRID systems". These technological infrastructures for scientific activity and research collaboration were briefly summarised in section 2.1.5 of the Conceptual Framework chapter.

5.1.6 Conclusions and directions for further research

In this section we have analysed the use of ICTs and particularly Internet technologies for the purposes of scientific communication and collaboration. The rapid evolution of Internet technologies and their world-wide availability to other social applications far from their initial focus on research communities, demands a better understanding in terms of the effects that might be produced within the science system. We are particularly interested in this research in the likely systemic influence of the use of these technologies for research communication and research collaboration.

First, we have tested the importance of these technologies for both remote and local research group collaboration. We found no significant difference in the use of these technologies for communication within the field of research under study compared with communication with members outside the research community. This provided empirical evidence for validating the theoretical hypothesis about "extended research groups" .

Secondly, we distinguished different patterns of use (in terms of both frequency and intensity) among this group of Internet technologies for different research activities. We found no significant association between the intensity of use of any of these technologies and the "seniority" of researchers or "research productivity". Other explanations of usage patterns might lie in the "learning process" discernable in the adoption of these electronic services. Moreover, the "multi-modality" of different technologies being used for different aspects of the research process is likely to support the need to embed patterns of ICT use in the collaboration context of scientific work and the formal/informal nature of the communication process. A discussion of a renewed model of scientific communication encompassing the digital infrastructure is provided in chapter 6.

Thirdly, we analysed the importance of these electronic networks (particularly the World Wide Web) as important infrastructures for the open and public dissemination of knowledge resources. Whether at the level of the individual researcher, or at research group level, research information (on projects, scientific publications and links to collaborators) is being distributed within these networks.

Finally, we analysed the need for more specialised and advanced computational resources and electronic infrastructures for scientific communication and collaboration. Empirical evidence on high-performance computing to support specialised research tasks - speech and language processing, access to large remote databases, access to print-servers and remote servers on speech and language documentation, or more demanding network architectures for remote collaboration - justify science policy initiatives to support the implementation of specialised scientific electronic services. In this last subsection, we briefly described the relatively recent focus of science policy on providing restricted national and international research networking infrastructures, environments for research collaboration ("*collaboratories*") and technological architectures for these electronic services such as *Grid systems*.

In section 5.2, we try to obtain a better understanding of the relevance of particular interactive and flexible environments for scientific communication, such as Newsgroups. We need to explain why these open and public electronic discussion fora, unlike the closed and restricted scientific networks described above (e.g. *National Research Networks*, *Collaboratories* and *Grid systems*), might constitute electronic environments for scientific and technical collaboration and "knowledge-exchange". In section 5.3, we extend our analysis of how open and public electronic networks are of fundamental importance for the distribution and dissemination of knowledge. Moreover, we analyse how the traditional non-electronic research networks and their inherent structures might be represented within electronic environments, allowing the discovery of digital knowledge bases.

5.2 Electronic "Invisible Colleges" and the Dynamics of Research Networks

5.2.1 Introduction

This section analyses interactive electronic systems supporting informal exchange of information. This research is focused on an empirical analysis of the scientific specialty of "speech technology". An extended analysis of the newsgroup activity of three specialised fora (comp.speech, comp.speech.research and comp.speech.users) is presented over the period 1992 - 2000. If we look at these discussion fora as electronic systems of social interaction, three particular research issues are of relevance.

First, some of these issues are related to the production of indicators characterising the activities of these newsgroups. How do these systems evolve over time or what is the growth function of these systems? Is this evolution similar to non-electronic diffusion processes or similar to the evolution of bibliometric results in the field of speech and language (as analysed in section 4.1)? What is the pattern of distribution of contributions by individual authors? Can we find similarities with the highly skewed distribution of productivity in publications or with the centrality and prominence of certain researchers (section 4.1)? What are the characteristics of the patterns of distribution of relevance of each discussion and the sectoral and geographical patterns of collaboration? Can we identify inter-sectoral and international collaboration groups similar to those encountered in collaborative research projects (see section 4.2)?

Secondly, these electronic social systems may or may not support the creation of social circles or "invisible colleges" within these scientific and technical communities. Of special importance is the identification of "who" is contributing to these electronic fora and networks. This analysis is conducted through a network analysis of these systems. It is important to understand whether these electronic collaboration structures are reproducing typical features of non-electronic research collaboration, as explained in sections 4.1 and 4.2, such as "invisible colleges", specialisation and division of labour and differences in connectivity and hierarchical structures.

In section 5.1 we analysed how researchers use different Internet technologies for various aspects of the research process. Here, we attempt to analyse how far Newsgroup systems are supporting the exchange of scientific knowledge, whether through the creation of new knowledge or through the distribution and use of existing knowledge. This leads to a content analysis of a sample of these newsgroup discussions. The empirical data have been thoroughly analysed and certain hypotheses have been tested.

5.2.2 Particularly Interactive Electronic Environments: Speech Newsgroups

The analysis proceeds by examining several interesting research issues, still open for discussion.

Operational Research Issues

1. What does the growth function of newsgroup activity resemble and what characterises the specialisation process in the generation of new sub-specialised newsgroups - is there a kind of "speciation" model of evolution?

2. What is the pattern of distribution of author's activity and prominence, on the one hand, and of the relevance of the discussion, on the other hand?

3. Should we not distinguish between "activity" indicators and "prominence" indicators of author participation and suggest an embryonic taxonomy for the web-metrics of newsgroups?

4. What is the collaboration pattern of these electronic discussions at the level of inter-individual collaboration and aggregate levels of inter-sectoral and inter-national collaboration?

5. How can one best characterise the social structure of these newsgroups and is the emergence of "electronic invisible colleges" apparent?

6. How far can these discussion fora go in the process of creating new knowledge, as opposed to the distribution and use of existing knowledge?

Hypotheses and Questions

H1) Different messages have different degrees of "relevance" and different authors also have different degrees of "activity" and "prominence". Does "Lotka's Law" of productivity apply here?

H2) There is likely to be a "Matthew effect" in newsgroup communities. The more active and "prominent" a contributor is, the more prominence he or she will receive.

H3) These electronic discussion fora support the electronic formation and evolution of "invisible colleges"- closely-knit, highly connected and bounded subgroups of authors.

H4) These electronic discussion fora help to create new knowledge, by allowing open discussion of new ideas and the reaching of a consensus on their validity and their scientific and technical standing.

These topics and hypotheses are empirically tested in the following sections.

5.2.3 The Evolution of Collaborative Electronic Environments and "Speciation"

Quantitative Indicators of Newsgroup Activity

This initial section covers the analysis of quantitative measures of newsgroup participation, with data from the three specialised speech newsgroups: comp.speech (from the starting date of the analysis - October 1992 - until April 1998); comp.speech.research (a newsgroup spawned by comp.speech dedicated to research issues, with data from May 1998, and analysis going until August 2000); and comp.speech.users (focused on application issues, and also with its roots in the initial discussion forum division, with data from May 1998 until August 2000).

One preliminary indicator of these newsgroup systems is the size of the electronic discussion forums. Table 5-2-I presents statistics for the three newsgroups, summarising overall size and activity.

Table 5-2-I - Size and Global Statistics for comp.speech, comp.speech.research and comp.speech.users

Descriptive Statistic	**Newsgroup**		
	Comp.speech.research (1998 - 2000)	**comp.speech.users (1998 - 2000)**	**comp.speech (1992 - 1998)**
Total number. of messages	3 304	4 870	20 011
Total different Discussions	1 462	1 733	11 130
Number of messages without reply	774	693	6 652
Total different Authors	1 505	1 646	8 783
Number of Authors with only 1 message	1 018	1 038	5 916
Number of Authors with only 1 Message (and Original)*	684	717	4 034

Source: Statistics computed by author's newsgroup data-mining application
* An Original message is a message that starts a new discussion in the Newsgroup.

As variables to measure the Size or Dimensions of the Newsgroup we could consider the following:

a) The *Total number of messages* exchanged within the newsgroup, during a certain time period;
b) The *Total number of different Discussion Threads* occurring in a Newsgroup, within a certain time period;
c) The *Total number of different Authors* contributing to a Newsgroup forum, during a certain time period

The statistics reported in Table 5-2-I provide a reasonable indication of the overall dimensions of a newsgroup and provoke the following observations:

- As the total number of different discussions is at least 50% of the total number of messages, it is apparent that many discussion threads are contained in messages to which no replies are made: - about 23% in the comp.speech.research, and 33% in the comp.speech newsgroups - but significantly less in the comp.speech.users (14%) - a fact that may be explained by the more "application oriented" nature of this last newsgroup.
- From the total number of different authors participating in the newsgroup (consistently about 50% of the total number of messages across groups), there is a very significant proportion of authors that make only one contribution (68% - comp.speech.research; 63% - comp.speech.users; and 67% in comp.speech). And of these single contributions, a major proportion of them *start a new discussion thread* but do not receive a reply (67% - comp.speech.research; 69% - comp.speech.users; and 68% in comp.speech). Explanations for this may include the non-personal nature of this medium, the low 'quality' of some contributions, or the (remote) possibility that a contribution is too brilliant to be understood by others.

From this initial descriptive analysis we can go on to discuss in the next section the pattern of evolution over time of these newsgroups, comparing this evolutionary process with theoretical expectations of "diffusion" processes, as well as characterising the process of sub-division of the initial newsgroup into two, more specialised newsgroups - which appears to be a general characteristic of electronic discussion fora and not just of these "speech" technology and science newsgroups.

The growth function of newsgroups and the "speciation" model of evolution

Newsgroups are electronic systems for social interaction in which participants contribute with new messages to introduce a new discussion thread or reply to an existing one. Thus, at any one point in time, several discussions are going on in parallel and it is likely that the relevance of the discussions is closely related to the "currentness" of the debate. On the other hand, as these electronic environments as well as full record of previous discussions, are permanently available, this represents a "knowledge repository" of all the ideas exchanged within these systems. This duality of the importance of information exchanged at a particular point in time as well as over the system's history highlights the significance of reaching a better understanding of the evolutionary nature of these systems. Furthermore, analysis of the newsgroup history reveals a natural process of specialisation or "division of labour", culminating in the later division of the initial system into two separate newsgroups. On the one hand, we are interested in the analysis of the particular time series of each of the newsgroups and the existence of potential similarities among these diferent sized systems ; on the other hand, we are particularly interested in obtaining a reasonable explanation for the "speciation" process.

A theoretical expectation for the evolution of these systems over time is its similarity with growth processes in general and diffusion processes in particular. If this were the case, a logistic growth curve would model reasonably well the growth of these systems over time. This assumption is tested as well as its relationship with the "speciation" process. For each of the three newsgroups, a monthly variable was computed, summarising the total number of messages discussed within that particular newsgroup in a given month. This enabled three different time-series data sets to be obtained, characterising the evolution over time of these newsgroups. In table 5-2-II the time-series data set for the largest newsgroup (comp-speech) is shown.

Table 5-2-II- Time-Series for comp.speech

	1992	1993	1994	1995	1996	1997	1998
Total Year	386	1 271	2 536	3 377	3 823	4 073	4 545
January		143	204	327	309	273	667
February		89	169	268	369	362	644
March		88	192	294	242	357	566
April		70	205	272	323	336	492
May		90	181	280	319	279	538
June		109	164	195	350	303	595
July		132	205	309	474	350	858
August		84	200	312	318	368	185
September		85	244	319	333	340	
October	209	116	261	240	345	367	
November	102	130	260	345	310	343	
December	75	135	251	216	131	395	
Cumulative Freq.	**386**	**1 657**	**4 193**	**7 570**	**11 393**	**15 466**	**20 011**

No significant patterns were found in the inter-monthly variation of the time-series data, which leads us to assume that regardless of the time of the year these electronic environments are likely to be used fairly constantly.

A trend analysis was conducted on this time-series, and two particular trend models were tested to fit the original data sets (the same analysis was conducted for the subsequent newsgroups - comp.speech.research and comp.speech.users). The respective graphs are presented in the Annex IX and show the growth functions of the three newsgroups. The discussion is presented in the following paragraphs.

A Quadratic trend model was compared to a logistic S-Shaped trend model in order to assess their "goodness of fit" with the original data sets.

A *logistic growth curve* is an S-shaped (sigmoidal) curve that can be used to model functions that increase gradually at first, more rapidly in the middle growth period and more slowly at the end, gradually levelling off towards a maximum value. The initial part of the curve is exponential; the rate of growth accelerates as it approaches the midpoint of the curve. At the midpoint ($K/2$), the growth rate begins to decelerate but the function continues to grow until it reaches an asymptote, K which is called the "carrying capacity" for the environment. This type of curve is frequently used to model biological growth patterns, both growth processes and diffusion processes, where there is an initial exponential growth period followed by a levelling off as more of the population is "infected" or some other factor limits further growth. The form of the symmetric logistic growth function is:

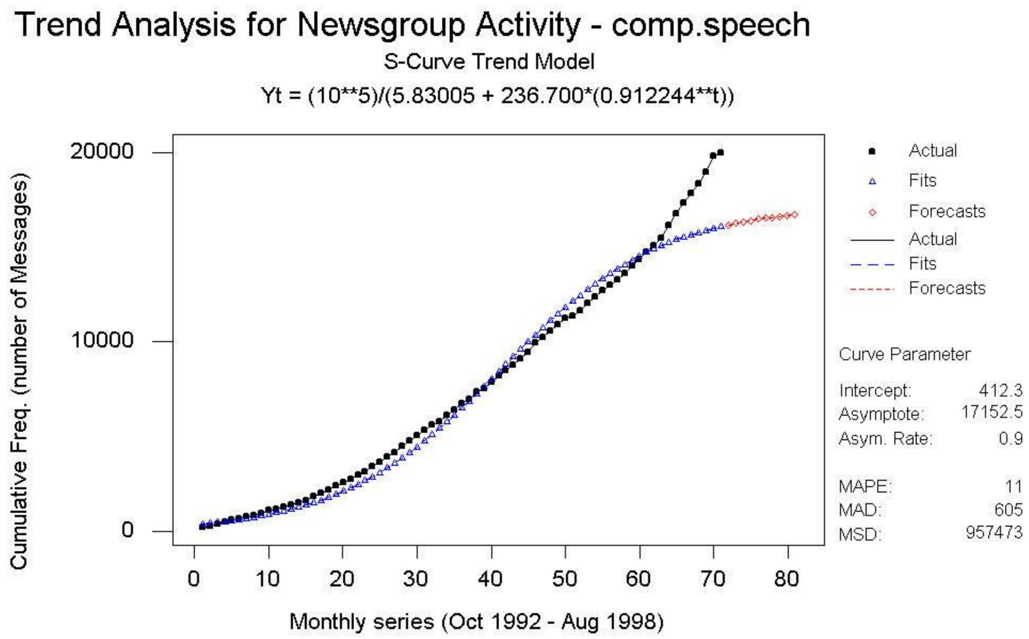
$y = K / (1 + \exp(a + b*x))$, where K, a, and b are parameters that shape and scale the function. The value of b is negative.

On the other hand, the quadratic growth curve is a simpler function usually reasonably well fitting with growth processes allowing for some "curvature" in the growth of the variable under analysis over time. This trend model is more applicable when there is a significant difference in the change of the variable every second period, as opposed to a percentage variation, where exponential models would be more convenient. The form of the quadratic growth function is $y = a + b*\mathrm{x} + c*x^2$, where a, b and c are parameters to be estimated and which scale the function behaviour.

The data results (see Annex IX) confirm the best fit of the quadratic growth function for all three newsgroups. In fact, this trend model performs much better than the logistic curve. This appears to contradict our initial theoretical assumption. But a closer look testifies that these electronic systems do not (as yet) have any *a priori* factor *limiting* the growth process, which would justify the "levelling off" effect typical of logistic S-shaped functions. This is probably the reason for the poor fit of the logistic model.

However, an alternative explanation may be related to the natural process of evolution of these systems in the long term, and specifically to the sub-division of newsgroups in separate discussion forums. As shown in Figure 5-2-1, the logistic model appears to fit well in the short to medium term evolution of the initial newsgroup (comp.speech). This is especially true **before** the mid-point of the logistic curve, from which the curve is supposed to inflect and start its ever decreasing growth rate.

Figure 5-2-1 - Logistic Growth Fitting to Original Dataset - comp-speech



However, if we fix the asymptote of the logistic curve - the limiting carrying capacity value - at a higher value (say 25,000) then we will have a mid-point value that is also higher, and the S-shaped curve would then have a better fit. This would require that the newsgroup would continue for a period longer than its real existence (terminating in August 1998). This fact is revealing of the impact of the division process and the special importance of "when" this process occurs.

One fact is testified by the empirical evidence: the "speciation" or sub-division process apparently (in this example, at least) occurs before any limiting factor constrains the growth of the initial newsgroup. As such, the factors explaining this process are not directly related to the natural growth or size of the electronic environment, but are likely to be related to other factors, such as the organisation of these newsgroups, or their "content" nature.

Another comment worth making is related to the similarity in the evolutionary nature of these newsgroups, especially when they are of comparative sizes. The comp.speech research and comp.speech.users, both resulting from the initial speciation process, have a similar pattern of growth. This is partly related to the similarity in evolution of these systems in the short term, as the two time-series cover a reasonably short period of about 2 years.

As a way of summarising the evolutionary nature of these electronic systems, two important characteristics are worth noting. In the first place, only in the medium to long run may we expect the growth function of these systems to exhibit a logistic shape. In other words, these systems need a long period of evolution before being constrained by any limiting capacity factor that would lead to a decay of their growing nature. This is quite understandable given the technological character of these systems, organised around relatively small discussion threads, as compared to the overall size of the system.

Secondly, the sub-division process or split of these electronic environments into two or more separate discussion fora occurs in a period of steady growth of the initial newsgroup. As such, the causal factors for this division are not related *per se* with the growth nature of these systems, but potentially with other organisational factors, such as the "epistemological" evolution of its contents, or the dynamics of the organisational structure of the initial system. This finding should be validated by comparison with other cases of specialised newsgroups. In this particular case, we have empirical evidence to justify the content nature of the newsgroup as the factor explaining the specialisation process (the emergence of a specialised forum for discussion of research issues - comp.speech.research; and a different discussion forum for application issues - comp-speech.users).

This specialisation process is remarkably similar to biological processes of evolution, and particularly the "speciation process", in which two groups of the same species occupy two different environments and, in adapting to them, evolve to eventually form two separate species. Hence, the use of the designation "speciation" to characterise the evolution of these electronic environments.

This argument brings us closer to the heterogeneous nature of these systems and particularly the variety or diversity characteristic of their organisation. On the one hand, this is related to the different nature of the actors contributing to the newsgroup discussions, and on the other hand to the diversity of "content" exchanged within these environments. A better understanding of the distribution of each participant's activity and discussion relevance is pursued in the next section.

5.2.4 Indicators of Activity and Prominence within Electronic Networks

Distribution of Author Productivity and Message Relevance - The Confirmation of Lotka's Law of Productivity

In the last section, the analysis of the evolutionary nature of these electronic environments pointed to the importance of organisational factors in explaining the behaviour of these systems. Among these organisational factors, two major characteristics are worth more thorough analysis: one is the identification of potential differences among the "actors" contributing to these networks of information interchange and in this section we will mainly focus on the quantitative measurement of these different patterns of participation. The second is the question of how heterogeneous and diverse is the content exchanged within these systems. Should we expect any particular form of distribution in the quantitative nature of the messages exchanged within these systems? These complementary dimensions for analysis are important both for the creation of indicators or quantitative measures of the activity and prominence of newsgroup participants (actor analysis) and message relevance (message analysis), as well as for the identification of patterns or structures emerging from these electronic interactions which will bring us closer to the second part of this section - the mapping of social structures.

For each of the three newsgroups, quantitative variables were computed for the measurement of an author's activity - the *identification of different authors*, the *total number of messages* contributed by each author, their *contribution over time*, and particularly the separation of *total number of original contributions* (starting a new discussion) from contributions replying to existing messages (*total number of reply messages*). Some identification variables were determined for characterising the authors (their *organisation*, the *type of organisation* - industry, university and other type - and the *country and region* of origin).

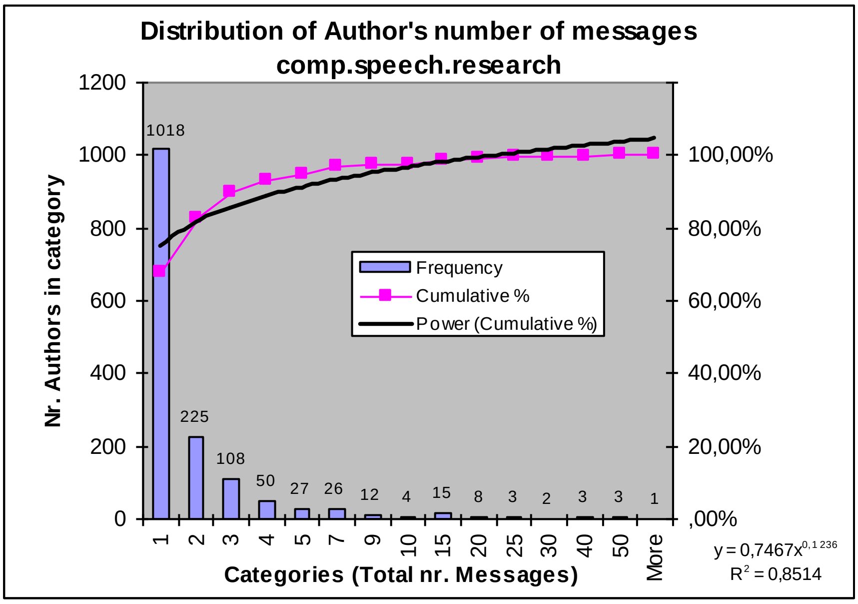
The same methodology was applied to the analysis of messages, variables being computed for the *identification of different discussion threads*, the *total number of messages* occurring in a discussion, the *identification of the original message* starting the discussion (as well as its contributor, the contributor's institution, institution type, country and region of origin), and the *total number of messages replying* to that discussion.

Let us focus first on the pattern of each author's productivity and then on the distribution of relevance of messages.

Author Productivity and the Confirmation of Lotka's Law

Newsgroups (the results are similar for all the three newsgroups) testify to a very heterogeneous pattern of author contributions. In fact, empirical evidence points to the "core-periphery" nature of the structure of these systems, at least with regard to the frequency of participation, given the high degree of skewness of the distribution function. The following graph outlines the distribution function of author contributions in the comp.speech.research newsgroup.

Figure 5-2-2 - Distribution Function of Author's Contributions



From the total number of 1,505 different authors (contributors) that were identified, a major proportion of them (67.6% - corresponding to 1,018 participants) contributed just one message. A small proportion, 39 authors, had contributed 10 or more messages each. During the period under analysis, only 1 author contributed more than 50 messages to the overall discussion of the newsgroup.

Theoretically this distribution conforms to an inverse power-law growth curve, and corroborates the pioneering work of Lotka in the 1930s, who analysed the distribution of scientists' published output. He found that a major proportion of the total output was derived from the work of just a few contributors, while the majority of the population contributed in a relatively insignificant way to total output. This pattern followed a consistent inverse power-law relationship. The same process is likely to occur in these electronic environments. A list of the most active authors within the newsgroup is presented in Table 5-2-3.

Table 5-2-III - List of Most Active Authors Participating in the comp.speech.research Newsgroup (May 1998 - Aug 2000) - Authors with 10 or More Contributions

AuthorID	Authoremail (Codified)	AuthorOrganisation	TotalMess	TotalOrigir	TotalReplie	OrganisationT	Country	Region
A7	A7@best.com	Bovik Research	174	18	156	Industry	US Domain Nar	United States
A61	A61@isip.msstate.edu	institute for signal and information pı	48	15	33	University	US Domain Nar	United States
A600	A600@micron.net	CyberGate, Inc.	45	15	30	Other Type	US Domain Nar	United States
A244	A244@ieee.org	Univ of Calif San Diego	44	6	38	University	US Domain Nar	United States
A78	A78@email.sps.mot.corr	Human Interface Technology Lab- P	39	2	37	Industry	US Domain Nar	United States
A746	A746@freepage.de_ANTI	FORWISS	37	5	32	Other Type	Germany	Europe
A121	A121@softsound.com	SoftSound	34	2	32	Industry	US Domain Nar	United States
A95	A95@sls.lcs.mit.edu	MIT Laboratory for Computer Scienc	28	1	27	University	US Domain Nar	United States
A54	A54@NOSPAM.trsvr.tr.u	Unisys	26	0	26	Industry	US Domain Nar	United States
A526	A526@Sunburn.Stanford.	Stanford University	22	4	18	University	US Domain Nar	United States
A111	A111@earthlink.net	EarthLink Network, Inc.	22	0	22	Other Type	US Domain Nar	United States
A233	A233@pacbell.net	Pacific Bell Internet Services	21	1	20	Other Type	US Domain Nar	United States
A151	A151@hotmail.com	Deja News - The Leader in Internet [	20	13	7	Industry	US Domain Nar	United States
A101	A101@world.std.com	The World Public Access UNIX, Bro	20	3	17	Industry	US Domain Nar	United States
A130	A130@fr.ibm.no-junk-mai	IBM France	20	0	20	Industry	France	Europe
A21	A21@zip.com.au	Zip Internet Access	20	0	20	Industry	Australia	Australia & NZ
A103	A103@speech.kth.se	Katholiek universiteit	19	17	2	University	Sweden	Europe
A680	A680@netfrog.net	USA Best Net	19	0	19	Other Type	US Domain Nar	United States
A162	A162@prosoniq.com	PROSONIQ PRODUCTS SOFTWAI	17	0	17	Industry	US Domain Nar	United States
A381	A381@ieee.org	The Hectic Eclectic	16	0	16	Other Type	US Domain Nar	United States
A140	A140@badran.freeserve.c	Customer of Planet Online	15	15	0	Industry	United Kingdom	Europe
A134	A134@watson.ibm.com	IBM T.J. Watson Research Center	14	2	12	Industry	US Domain Nar	United States
A57	A57@fr.ibm.no-spam.cor	IBM France Speech Recognition	14	1	13	Industry	France	Europe
A339	A339@fri.uni-lj.si	RCU, University of Ljubljana, Sloven	13	9	4	University	Slovenia	Europe
A35	A35@imt.unine.ch	Institut de Microtechnique - Universi	13	5	8	University	Greece	Europe
A210	A210@fr.ibm.NO-JUNK-M	IBM France	13	1	12	Industry	France	Europe
A202	A202@bestweb.net	BestWeb (bestweb.net)	13	0	13	Other Type	US Domain Nar	United States
A787	A787@scl.ece.ucsb.edu	University of California, Santa Barba	12	1	11	University	US Domain Nar	United States
A415	A415@my-deja.com	Deja.com - Share what you know. L	12	0	12	Industry	US Domain Nar	United States
A116	A116@clear.net.nz	Customer of Telecom Internet Servic	11	10	1	Industry	New Zealand	Australia & NZ
A11	A11@cui.unige.ch	University of Geneva	11	10	1	University	Switzerland	Europe
A714	A714@altavista.net	Deja News - The Leader in Internet [	11	9	2	Other Type	US Domain Nar	United States
A532	A532@cs.joensuu.fi	University of Joensuu / Dept. of Corr	11	8	3	University	Finland	Europe
A628	A628@fr.ibm.com.invalid	http://www.remarq.com: The World's	11	1	10	Industry	US Domain Nar	United States
A1253	A1253@email.mot.com	Motorola SSTG	11	0	11	Industry	US Domain Nar	United States
A866	A866@kgw.tu-berlin.de	TU-Berlin	10	2	8	University	Germany	Europe
A395	A395@anu.edu.au	RSISE	10	1	9	University	Australia	Australia & NZ
A979	A979@netwood.net	Posted via Supernews, http://www.s	10	0	10	Other Type	US Domain Nar	United States
A711	A711@iowegian.com	Iowegian International Corporation	10	0	10	Industry	Greece	Europe

Source: Author's spreadsheet for Author Analysis, after results from data-mining newsgroup application
Note: The ID of the Authors, as well as their e-mails, are codified for confidentiality reasons.

Two points should be highlighted from the above analysis. Firstly, we should attempt to distinguish "activity" measures of contributions from "prominence" indicators. In fact, even among the most "active" contributors there are differences in their patterns of participation - see the discussion below. Secondly, as is shown in table 5.2.III, the sectoral pattern of participation as well as the international character of these electronic environments are clear. A more detailed comment is provided in the section summarising the sectoral and international patterns of collaboration.

With regard to patterns of "activity" of different contributors we can make the following distinctions shown in the following table, although one should be a little cautious of interpreting them too literally.

Table 5-2-IV - Taxonomy for ACTIVITY Indicators Within Newsgroups

ACTIVE Contributors	**Permanent Contributors** - This group combines the characteristics of both the following groups. These actors both contribute to starting new discussions, and reply to on-going discussions. They consistently participate in the newsgroup discussions.
	Original Contributors - This group of actors usually participates by contributing original messages to the newsgroup, starting new discussion threads
	Frequent Participants - This group is built around participants who often contribute to the discussions, replying to already existing discussion threads, but rarely if ever start new discussions.
Threshold level of Activity	
NON-ACTIVE Contributors	**Rare Participants** - This group of participants reveals a low level of involvement in the newsgroup discussions - the major proportion of this group contributes only 1 or 2 messages.
	Observers - This group of observers does not actually participate in the newsgroup discussions. They just use these electronic environments to search for information, without actively participating.

We cannot directly measure the size of the "observers" group by an analysis of archival records of newsgroup messages. On the other hand, the size of this group reveals how important these electronic environments are in terms of being repositories of information and knowledge and potentially "knowledge bases" to be used for systematic investigation.

The threshold level distinguishing the NON-ACTIVE group from the ACTIVE groups should be fixed based on individual researchers' assumptions supporting the overall investigation or the distribution function of participation in the newsgroup under analysis. The empirical investigation of these three newsgroups would suggest that a level of 5% of the overall population for distinguishing the most active contributors is a realistic value for fixing this threshold level.

Similarly, some kind of measure should help to distinguish between the different categories within the ACTIVE group. This can perhaps be based on a simple index weighing the difference between original messages and reply messages against the total number of messages for a certain author. This index should not exceed any level defined by *a priori* theoretical assumptions. Let us suggest a 10% to 20% confidence interval. This index would take the form of

$IActiv_a = (OrigMessgs_a - ReplyMessgs_a) / TotalMessgs_a$, where

$OrigMessgs_a$ is the total number of original messages by author a, $ReplyMessgs_a$ is the total number of reply messages by author a, and $TotalMessgs_a$ is the total number of messages contributed by author a. If the value of this index is negative and lower than (0.50 + the confidence interval value), then this author is part of the frequent participants group. If the value of this index is positive and greater than 0.50 + the confidence interval value, then this author is part of the original contributors group. In any other case, the author is part of the Permanent Contributors group.

Let us assume that we have fixed the threshold level for distinguishing the ACTIVE contributors at the demanding level of 2.5% of the total population of actors. The most active 2.5 % of authors would correspond roughly to the 39 authors listed in Table 5.2.III. From these select few, we could distinguish three different groups (permanent contributors, original contributors and frequent participants). For this we could establish a 20% confidence interval for the Index of Activity (IActiv) of the 39 authors.

This would result in the following three sets:
Permanent Contributors {A61, A600, A526, A101, A35, A866, A339, A714, A532}
Original Contributors {A151, A103, A116, A11, A140}
Frequent Participants {A7, A244, A78, A746, A121, A95, A54, A111, A233, A130, A21, A680, A162, A381, A134, A57, A210, A202, A787, A415, A628, A1253, A395, A979, A711}

Furthermore, the activity behaviour of authors should be complemented by a detailed analysis of the "prominence" character of their participation within the newsgroup. This measure is closely linked with the participation of an author in *more* or *less* relevant discussions (the relevance being measured by the frequency of replies to a given discussion).

We should start the distinction between activity indicators and prominence indicators with the following statement: to be an ACTIVE author is a necessary but not a sufficient condition to become a PROMINENT author. In fact, one first has to be active in the discussion fora to be "recognised" and progress from that to increase the "prominence" awarded by other contributors.

We would suggest three different degrees for the prominence indicators of an author contributing to these newsgroups.

Table 5-2-V - Taxonomy for Definition of Prominence of Authors Contributing to Newsgroups

Very Strong Definition	One has to comply simultaneously with both conditions of prominence: weak and strong definitions of prominence - see the following categories.
Strong Definition	To become a "prominent" author one has to have a high level of original activity (to be an "original contributor"), and simultaneously a high level of reply to the author's original contributions.
Weak Definition	To become a "prominent" author, one has to participate in Newsgroup discussions with a high level of replies (highly relevant discussions)

The thorough verification of these "prominence" conditions should follow from the analysis of the pattern of the distribution of message's relevance.

Discussion's Relevance

It has been hypothesised that "different messages have different degrees of 'relevance' and different authors also have different degrees of 'activity' and 'prominence' ". This latter component we have covered in the previous paragraphs. We now turn to the distribution of the "relevance" of the discussions occurring in a Newsgroup.

As a first approximation we use the term "relevance" as a quantitative measure of the total number of messages which in combination form a given discussion. From this quantitative perspective, it is reasonable to say that the more relevant a discussion, the more intensive it is and, therefore, the higher will be the number of messages constituting that discussion thread. A more qualitative assessment of the "content relevance" of the discussions is provided in section 5.2.7, concerning the content analysis of Newsgroup messages.

The three newsgroups were analysed in terms of the distribution of the contributed messages as well as the resulting discussion threads. The results for the comp.speech.research newsgroup are discussed in this section. A more comprehensive discussion of the results for the other newsgroups is beyond the scope of this investigation.

The following figure and table present the results for discussion.

Figure 5-2-3 - Distribution of the Discussion's Relevance - comp.speech.research Newsgroup

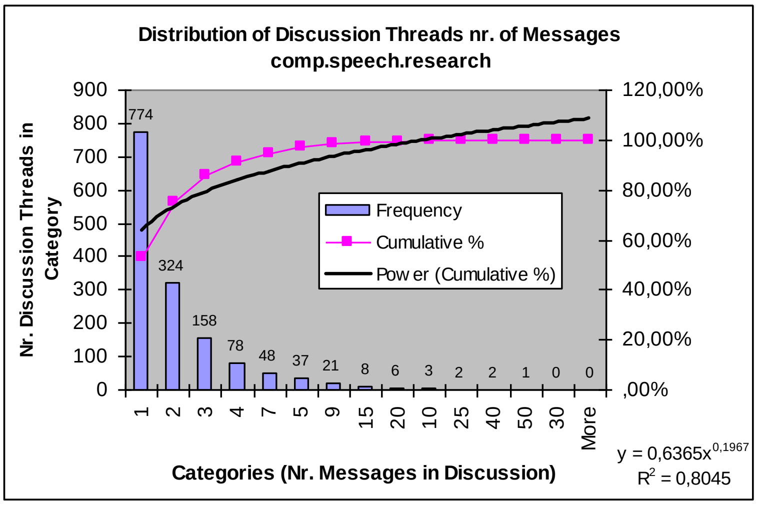


Table 5-2-VI -List of the Most "Relevant" Discussions within comp.speech.research Newsgroup (May 1998 - Aug. 2000) - Discussion Threads with 10 or More Messages

Discuss	Subject	TotalMess	TotalOri	TotalRep	Author1OrgT	CountryAuthor
126	dragon nat. speak.... l&h voice xpress, ibm viavoice eval	45	1	44	Industry	US Domain Nam
1361	Future Shock, an understatement!	37	1	36	Industry	Germany
904	Non Power-of-2 length FFT	36	1	35	Industry	US Domain Nam
1192	Quest for a low noise PC	24	1	23	Other Type	US Domain Nam
905	Why so much FFT and so little Wavelet?	24	1	23	Other Type	Denmark
602	Justification for SNR equation using variance	20	1	19	Industry	US Domain Nam
1235	This can't be right	19	1	18	University	US Domain Nam
210	Q: Free Speech Synthesizer Software?	19	1	18	Industry	United Kingdom
741	Pitch detection algorithm	17	1	16	University	Germany
508	Machine Demonstrates Superhuman Speech Recognition A	16	1	15	Other Type	US Domain Nam
57	Different definitions of cepstra?	16	1	15	Industry	US Domain Nam
80	Via Voice crashes during training	15	1	14	Industry	US Domain Nam
412	Linux speech recognition	14	1	13	Other Type	US Domain Nam
311	Speech synth which can sing?	14	1	13	Industry	US Domain Nam
1160	US:GA: DSP Opening-Atlanta	13	6	7	Industry	US Domain Nam
53	Tone replicator circuit... ?	13	1	12	Other Type	US Domain Nam
1428	Thanks for FAQ info!	11	1	10	Other Type	France
1413	How do I do this?	11	1	10	Other Type	US Domain Nam
262	Talking about HMM scores...	11	1	10	University	Germany
1295	Information on speech synthesis?	10	1	9	Industry	US Domain Nam
1135	Announcing the Microsoft Speech SDK 5.0 Beta	10	1	9	Industry	US Domain Nam
56	Evaluation procedures for speech recognition	10	1	9	Industry	Norway

Source: Author's spreadsheet for Message's Analysis, after results from data mining newsgroup application

The above results confirm the existence of a highly skewed distribution of the relevance of discussions within the newsgroup (this is consistent throughout the three newsgroups analysed). From the total of 1,462 different discussions, a major proportion (52.9% corresponding to 774 discussion threads) were classified as single messages (discussions with only one message and no reply). Just a relatively minor percentage (2.9% of the total number of discussions) were grouped under the categories of seven or more messages per discussion. Only three discussions involved more than 30 different messages and only 22 discussions had 10 or more messages (see table 5.2.VI).

This is empirical evidence corroborating the hypothesis that "discussions have a widely varying degree of relevance" at least with regard to the intensity of the discussions exchanged within the newsgroups. Theoretically this sounds a note of caution for assessing the intensity of the "interactive" nature of these electronic environments, as we can see that this is not equally valid for the whole set of information posted and exchanged within these newsgroups. However, we are not able to say anything about the "repository effect" of these discussions, as we can measure the number of participants in each discussion but can say nothing similar about the frequency with which these discussions have been "observed" by others.

The more interesting findings result from a combination of both analyses - the distribution of author's activity and prominence, on the one hand, and the analysis of message relevance, on the other. A detailed analysis of the three newsgroups allows the conclusion to be made that the most active authors are also those participating in the most relevant discussions (this is empirical confirmation of the *weak definition of prominence*). From the most relevant discussions in each of the newsgroups, a significant proportion of them included the authors that are the most active in the respective newsgroup (77.5% of the cases in comp.speech.research; 92.5% of the cases in comp.speech.users; and 65% of the most relevant discussions in comp.speech). Secondly, in terms of the most relevant discussions in each of the newsgroups, a significant proportion of them were originated by the same active and weakly prominent authors, which constitutes empirical evidence for the *strong definition of prominence* (47.5% of the discussions in comp.speech.research; 45% of the cases in comp.speech.users; and significantly less in comp.speech - 25% of the total 40 cases).

The following table summarises the results of this analysis for the 40 most relevant discussion threads for each newsgroup.

Table 5-2-VII - Percentage of Discussions with Most Active Authors

Percentage of Discussions	Comp.speech.research	Comp.speech.users	Comp.speech
Authors among the list of most active in these relevant discussions *	77.5%	92.5%	65.0%
Authors among the list of most active who initiate these relevant discussions	47.5 %	45.0%	25.0%

Note: * If there were at least 3 authors participating in the discussion and belonging to the list of most active, the discussion was considered for the overall percentage of this category.

This leads to a quasi-confirmation of the hypothesis about the existence of a "Matthew effect" in these electronic environments. In fact, the more active a contributor, the more prominent he is likely to become and, if he participates in relevant discussions, there is a reinforcing mechanism which increases his "visibility" and prominence. What we have done here confirms this, by analysing the result of this "Matthew effect virtuous circle", selecting the most relevant discussions, and determining who were the authors posting these messages and particularly the ones initiating the discussions.

Full confirmation of this hypothesis would require a close longitudinal analysis of each author's prominence, which is beyond the scope of this study. However, as the initial newsgroup (comp.speech - 1992 - 1998) reveals a less clear "prominence" measure for the most active authors, and the two subsequent newsgroups (comp.speech.research and comp-speech.users, in the period 1998 - 2000) reveal a stronger measure for the prominence indicator, we could tentatively hypothesise that a more thorough longitudinal analysis would corroborate the above results. Additional evidence is provided by the fact that the most active authors in the initial newsgroup (comp.speech) are also the most active ones in the subsequent separate newsgroups (some constitute the most active in the research forum, others in the applications forum). It seems, therefore, that some kind of prominence effect is occurring during the "speciation process" and consequently over time.

This analysis is taken a stage further in the section on the network analysis of these newsgroups. In the next section, however, we focus on the sectoral and international collaboration patterns exhibited in these electronic environments.

5.2.5 Inter-sectoral and Inter-national Electronic Collaboration

Computer networks and electronic environments are commonly believed to facilitate remote collaboration. Indeed, their technological capabilities usually facilitate the connection of geographically dispersed collaborators in an easy and effective way. Previous research has demonstrated the usefulness of these technologies for scientific collaboration over an international environment (Walsh and Bayma, 1996). Moreover, science and technology policy is usually interested in fostering international collaboration of scientists and practitioners in order that resources can be shared and a more effective division of labour in the whole research effort achieved with duplication of efforts being avoided. We would, therefore, expect these particularly interactive environments - Newsgroups - to reflect the international character of electronic communication. The speech science and technology community involves participants from different research 'sectors' (universities, industrial laboratories, government research centres and other institutions). The nature of collaboration should reflect this inter-sectoral involvement. Both these dimensions - the international character of the collaboration as well as the inter-sectoral nature of the interchange of information - are assessed in this section.

A special procedure was developed in order to assess the international and sectoral nature of the contributors to the newsgroup.[34] For each newsgroup, the 10% most active contributors were selected and their nation and sector of origin were examined. A second analysis was made of the 40 most relevant discussions in each newsgroup which establishedthe proportion of cases involving international and inter-sectoral collaboration.

Table 5-2-VIII - International and Sectoral Pattern of Collaboration (40 Most Relevant Discussions)

Percentage of Cases	Comp.speech.research	Comp.speech.users	Comp.speech
International collaboration	70.0%	50.0%	82.5%
Inter-sectoral collaboration	60.0%	50.0%	70.0%

Note: If the discussion involved more than three participants belonging to different countries of origin or different organisational sectors, these discussions were computed as positive in the overall percentage of international collaboration or inter-sectoral collaboration, respectively.

Table 5-2-VIII shows a summary of the statistics on international and inter-sectoral "collaboration" (defined as three or more participants belonging to different countries or sectors) in the 40 most active discussions for the three newsgroups. It indicates the following:

- The level of international collaboration is significant, being relevant in more than 80% of the total number of discussions (83% of the comp.speech newsgroup) and 50% in the least significant case (the comp.speech.users discussion forum);

- The same is true for the inter-sectoral pattern of collaboration within these newsgroups (50% of the discussions in the comp.speech.users forum were categorised as inter-

[34] The raw data files (newsgroup messages) do not contain specific information about the nationality of the contributors or the type of organisations to which they belong. It was therefore necessary to create an approximate indicator for these variables. The *country* and *region of origin* fields of the contributors were created based on the e-mail of the author field (a filtering algorithm was programmed for assessing the domain name of the e-mail address – either a country level domain name or the domain name -.org, .com, .edu, … which were attributed to the US). Some inconsistency in the final results may be due to the existence of institutions with a .com domain name which was not US origin. The region field was derived from the country field. Six regions were defined: United States, Canada, Europe, Australia & New Zealand, Japan, Asia and Pacific. The *organisation type* field was created based on the organisation of author's field in combination with the author's email address. Three categories were considered: University, Industry, and Other Type. Distinctions between type were based on content analysis. These automated procedures have been confirmed by visual inspection and manual verification of the most active contributors and the most relevant discussions, which demonstrated their robustness.

sectoral, and this percentage is higher for the other two newsgroups - 60% in comp.speech.research and 70% in the comp.speech newsgroup);

- The comp.speech.users newsgroup reveals both a lower level of international collaboration as well as inter-sectoral collaboration. This is in accordance with the more "applied" nature of this newsgroup, as most of the intra-sectoral discussions were Industry-Industry exchanges of information, and the more intensive participation of US contributors in this forum, which explains a higher concentration of the discussions in the United States.

One of the most important findings from these newsgroup data sets is the significant level of participation of university contributors. This contrasts with previous research on "cold fusion" (Lewenstein, 1995), which found little participation from active researchers. The higher level of participation may be related to the natural evolution of these electronic environments (since seven years have elapsed from that particular investigation), or to a specific characteristic of this particular community.

Further information on the pattern of international and organisational type can be derived from the list of different authors making at least two contributions. The top percentile of the most active contributors was selected, and their international and sectoral characterisation was inspected.

Table 5-2-IX - International and Organisational Origin of 10 % Most Active Contributors *

Frequency of Cases	Comp.speech.research Total Cases = 49	Comp.speech.users Total Cases = 61	Comp.speech Total Cases = 287
International origin			
US Domain name	31	54	188
Canada	1	5	9
United Kingdom	1		32
Ireland			2
Germany	2		10
France	4		9
Belgium		1	6
Netherlands			1
Sweden	1		4
Switzerland	1		3
Austria			3
Greece	2		
Portugal			1
Finland	1		1
Norway	1	1	1
Slovenia	1		
Romani			1
Russia			2
Australia	2		8
New Zealand	1		
Japan			2
China			1
Hong Kong			1
Korea			2
Organisational origin			
Industry	21	39	111
University	15	5	103
Other Type **	13	17	73

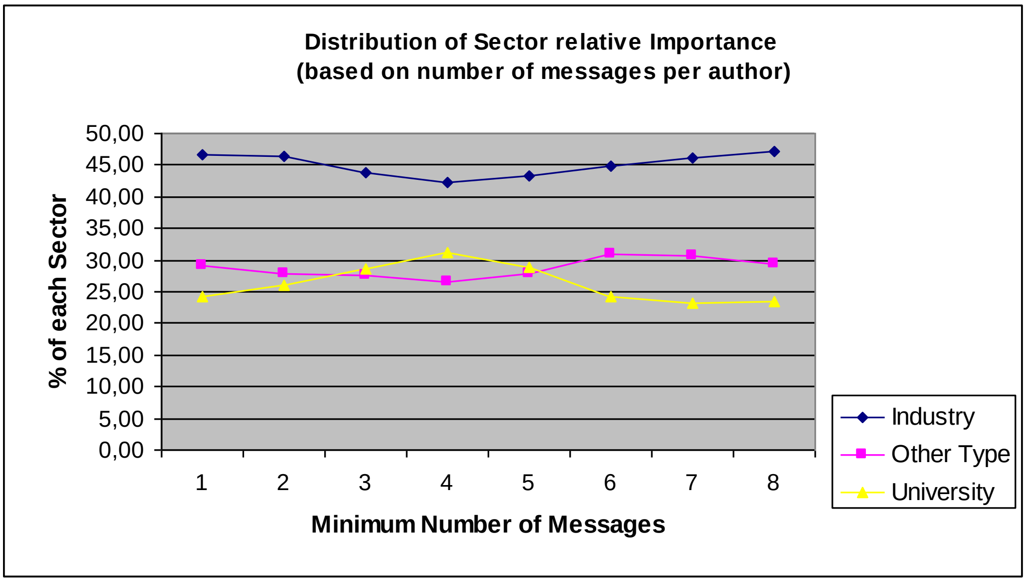
Note: * The three samples of 10% most active contributors within each newsgroup were taken from the total population of different authors excluding the contributors with a single message.
** Within this category we may have a combination of organisation types other than industry and university but also some false negatives of those two categories, which, for computational reasons, were not properly categorised. Therefore, some cases currently in this category may really belong to one of the other categories.

As might be expected, the proportion of US participants is significantly higher than the whole of the rest of the international contributors (even allowing for a probable over-enumeration of the US Domain names category, given the computational procedure used to determine international origin of contributors). However, the diversity of countries to which the “most active” participants belong is also of note. International collaboration spans almost all parts of the developed world from the US to Australia and New Zealand, and includes a significant number of European countries. This is most notable in the comp.speech.research and comp.speech newsgroups and less so in the comp.speech.users discussion forum, where the concentration on the US is very strong.

With regard to the organisational sector of origin of the most active contributors, both the University and Industry realms are well represented in these electronic environments - the strong participation of the University sector in these computer networks confirming the results discussed in previous paragraphs. Only in the comp.speech.users application-oriented discussion forum is Univeristy participation not so noticeable (but, based on the subject matter of this forum, this result could theoretically be expected).

However important the above results, some caution is needed in the detailed assessment of the relative importance of the sectors for the activity of the newsgroup. The following graph of a detailed analysis of the whole set of discussions exchanged in the comp.speech.research newsgroup reveals some significant differences in the relative weight of the different sectors when the distribution of the number of messages per author is analysed.

Figure 5-2-4 - Relative Importance of Each Sector when Compared to the Activity of Participants



When the full set of 1,505 different authors contributing to the newsgroup is analysed, the relative weight of the University contributors is higher in the interval of authors with two to six messages each, than in the extremes. And, conversely, the relative importance of contributors from the Industry sector is higher at the extremes of the scale (either participants with just a few messages - 1-2, or very active contributors - more than eight messages each).

These final results point to the need to conduct a more thorough analysis of the full set of messages, as well as the full set of contributors, in order to assess rigorously the pattern of international and sectoral characteristics in these electronic environments. On the other hand, they also demonstrate the need for these environments to provide a better categorisation of the ethnographic characteristics of their contributors. This more rigorous analysis is beyond the scope of this study.

In previous sections we analysed mainly the quantitative nature of these electronic systems. In the first place, we were interested in the evolutionary pattern of the newsgroups. The growth function of these systems was analysed and we assessed the applicability of theoretical distributions typical of growth and diffusion processes, namely the logistic model. We also sought an explanation for the "speciation" process occurring in the evolution of these newsgroups and supporting the sub-division and creation of more specialised discussion fora. Secondly, we studied the patterns of distribution of author activity and prominence on the one hand, and the distribution of each message's relevance on the other. We confirmed the heterogeneity and diversity of the discussions occurring within these environments and among authors contributing to the discussions. Moreover, some indicators of activity, and a taxonomy for contributors to newsgroups, were suggested. The prominence of contributors was found to be related to the relevance of the discussions and also with the level of activity of the contributors over time. Finally, we analysed how far these systems support international and inter-sectoral collaboration, as environments facilitating remote collaboration.

In the next part of this section our attention will focus on the social structure of newsgroups, with particular emphasis on detection of patterns of relations, as measured by the interchange of information over time. Since we have identified the relatively active contribution of some participants and confirmed the international and inter-sectoral collaboration nature of these contributions, then, over time, some kind of social structure is likely to emerge. We test the hypothesis of emerging "invisible colleges" or social circles of collaborators, who, by repeatedly exchanging information in different discussions over time, get to know each other and come to form an "invisible" social structure of collaboration, like those of the natural philosophers of seventeenth century England, or their contemporary scientific peers, who informally exchange information about their current research in order to validate their knowledge claims.

In the next section, we will focus on the assessment of the potential existence of such a social structure. In the final part of this section (subsection 5.2.7), we will assess what kind of knowledge (if any) is being exchanged within these electronic environments.

5.2.6 Structures of Collaboration and Electronic "Invisible Colleges"

We have seen how contributors to these newsgroups are heterogeneous, both in terms of their activity and the prominence they demonstrate through interacting within these electronic discussion fora, as well as in terms of particular characteristics such as their national origin or organisation. In this section, we are more interested in the activity of "groups" of contributors (scientists, practitioners or other participants) and the potential emergence of social structures, than in the individual activities of single authors. In particular, we examine the "relationship" of each author with all other authors within these electronic environments. The main focus will be on the social networks that these electronic systems may help to support and their evolution.

Using this relational perspective, each contributor to a newsgroup should be considered as an Actor, whose actions (in this case, a contribution in the form of messages to the newsgroup) take place in a complex web of relations with other actors. The specific relation we are measuring is the *exchange of information and knowledge* and we are also interested in measuring this relation over time, as we have a longitudinal data set – i.e. looking at the whole set of messages exchanged within these newsgroups over time.

We are interested in assessing the degree to which these computer networks and electronic environments support the activity of social networks. Furthermore, we want to test the hypothesis that, over time, these computerised social networks come to exhibit an "invisible college" or group-like structure. It is expected that these groups of actors/participants will have a higher degree of connectivity and will be tightly bound together. Thus, we need to conduct a structural analysis of these electronic systems and, specifically, to perform an analysis of the cohesiveness of subgroups in these networks. If this structure can be validated, then these electronic systems (especially tailored to remote collaboration and to providing effective tools for distance communication) may also support the usual informal communication function in scientific and technical communication systems. As such, they may indeed represent a significant step forward in the extension of the traditional scientific communication paradigm into electronic environments.

Let us now take a closer look at the methodology followed to conduct the structural analysis.

- In the first stage, we need to establish a boundary for our network analysis, selecting the top percentile of the most intensive discussions that occurred in each of the three newsgroups. This initial matrix contains two sets of information: the discussion threads and the corresponding array of authors contributing to each thread. From this data set we are able to compute a co-author matrix representing the relationship of each author with every other author in the network;
- Secondly, we attempt to determine the subgroup structure of the networks (one network for each newsgroup). From the initial co-authorship matrix we apply some techniques in order to "aggregate" the most connected authors with one another. Moreover, we rigorously specify conditions for the degree to which these subgroups should be connected, conducting a "clique" analysis of the data. Potential "cliques" - subgroups of authors who are tightly bounded and highly connected - are identified;
- In the third stage, we conduct a centrality analysis determining two complementary indicators: the *degree of centrality* of each author - an indicator of the intensity or Activity of each member of the network; and a *flow-betweenness centrality* - a measure of the "importance" of each author in the overall connectedenness of the network, i.e. how important each author is in bridging the different subgroups within the network.
- Finally, we compare the subgroup structure of the three different newsgroups in order to assess any potential structural evolutionary pattern from the initial newsgroup (comp.speech) to the resulting specialised newsgroups (comp.speech.research and comp.speech.users).

The analysis is conducted for the three social networks represented by the three newsgroups. The peculiarity of these three networks is their evolutionary relationship. As noted previously, one of these networks (linked to the comp.speech newsgroup, running from October 1992 until August 1998) preceded the other two consequent networks (linked to the comp.speech.research and comp.speech.users newsgroups, both running for the period May 1998 - August 2000). Consequently, in reality we have two main objectives: 1) the identification of cohesive subgroups in each of these networks; 2) potential linkages between the three newsgroup structures that may further inform us about the evolution of the social networks to which they are linked.

We are analysing a specific kind of social network, designated as an Affiliation Network. The peculiarity of such networks is that we do not really have network data characterising the relationship of every Actor involved in the network and the quantitative measure of the relationship with all the other actors in the same network. What we do have is the identification in time of a specific Event, in which different Actors simultaneously participate. In this particular case, we have Discussion Threads, in which scientists, technicians or other practitioners contribute with messages (information). This kind of network is a particular case of a two-dimensional network, in which we have two different Sets (a set of Events and a set of Actors participating in those events). The simpler, more familiar, kind of social networks are typically one-dimensional networks in which we measure a relation between a single set of actors.

The first step in the network analysis is to gather the Affiliation network data (list of Events, in this instance Newsgroup discussions, and corresponding Actors – the participants on those discussions). This was computed for the top percentile, most relevant discussions, corresponding to 68 discussions in comp.speech.research. The second step in the network analysis is the computation of a one-mode network based on this Affiliation network. From the affiliation data, we are able to compute the co-membership matrix of this network. There were 265 different authors participating in the network.

Based on the affiliation data of authors participating in each discussion, we can identify the relationship or links for each author of a message with every other author. In this case, we have a valued relation, as it is not only important to identify whether a certain author has links with another author (binary relation), but also how strong those links are in terms of the frequency of interaction (value relation). We will thus be able to identify the effective number of times a certain author has "collaborated" or has been involved in discussions with other authors. Using this information we can derive a co-membership matrix in which the diagonal values indicate the total number of contributions for each author and the off-diagonal values measure the intensity of the relation of each author with every other author.

A first approximation for detecting "group" structure within the network consisted of a core-periphery analysis to determine a subset of the initial co-membership matrix based on the interactions between the "core" members of the network. In the comp.speech.research newsgroup this subset of "core" member's consists of only 88 actors who, by the frequency of their interaction with each other, reveal a level of connectedness distinctly above that in other sub-groups. The computational algorithm (using the UCINET V network analysis package) for selecting these core authors is based on a series of permutations of the rows and columns of the initial matrix in order to identify the most interactive groups. The approximate number of actors is defined *ex-ante* by theoretical expectations. From this subset of the co-membership matrix, we could already detect some "group" structure, in the sense of blocks of actors with higher degrees of interaction within the group than outside the network.

A more detailed analysis involves the identification of the subgroup structure by selecting the "cliques" of authors under more or less restrictive conditions. A "clique" is a group of Actors that interact only with other members of that clique without exceptions. Different sets of cliques may be defined by varying the number of interactions required to form a clique. Based on 10 interactions among (and only with) other members of the clique there are 21 cliques in the comp.speech.research newsgroup while based on 20 interactions, 10 cliques can be defined. A clique is clearly one form of "electronic invisible college".

The clique analysis provides empirical evidence for confirming our hypothesis that these electronic environments support the creation of electronic "invisible colleges".

This conclusion is supported by analysis of the resulting "cliques" or subgroups in the two other newsgroups. From the initial 104 most relevant Discussions in the comp.speech.users newsgroup, we were able to identify a network of 268 members. From the interactions of these members, and applying the same restrictive conditions for the intensity of collaboration within the members of the network, we could identify 31 "cliques" or subgroups with a level of interaction of at least 10 discussions, and a total of 11 subgroups when we increased that cut-off value to 20 interactions.

In the comp.speech newsgroup a comparable result was obtained. From the 380 most intensive discussions, we could build a network of 1,534 members. The cohesive subgroup analysis of this network revealed the existence of 75 tightly bounded subgroups when the level of interaction among the members of groups was fixed at 10 links and a total number of 19 subgroups when one is more restrictive and demands an intensity of at least 20 interactions between the members of these "invisible colleges".

Building upon the results of the subgroup analysis, we are also interested in the differentiation of members with respect to their visibility or "centrality" within the network. Even among the members of these very interacting subgroups, we would expect to have hierarchies from the most "central" to the less visible Actors.

A common measure of "centrality" is given by the visibility of actors within the network. It is commonly assumed that the more "Active" an actor is, the more visible he or she will be within the network. As such, the most Active Actors would have a proportionally higher number of links with other Actors, and consequently be more "visible" and "central" within the network. We use the degree of centrality measure to assess the centrality of each author in the whole network. The most central Actors, identified by the structural analysis, correspond largely to the most Active Actors identified previously,[35] where the distribution pattern of authors' activity was analysed. Furthermore, the most central Actors in terms of flow-betweenness (connecting or bridging two separate groups in the network) are very likely to be the authors with the greatest "prominence".

Structural Evolution of Newsgroups and "Division of Labour"

Finally, we are interested in determining any potential linkages between the evolution of the structures of these three networks. It is worthwhile to validate two assumptions. The first is that the structures of the two specialised newsgroups are different. In other words, the memberships in these networks, or evolving "invisible colleges" of intensive communication, are distinct to a certain degree. This would indicate the different nature of the two newsgroups, in terms of their application and research specialities, respectively. The second assumption is that from the emerging structures of the initial newsgroup (in the period 1992 - 1998) some kind of structural evolution would lead to the maintenance of some of these structural patterns for each of the resulting separate networks - reinforcing the hypothesis of a "speciation" process of evolution.

Table 5-2-X- Comparison of "Clique Structures" for Groups with Intensity of Interaction Greater than or Equal to 20

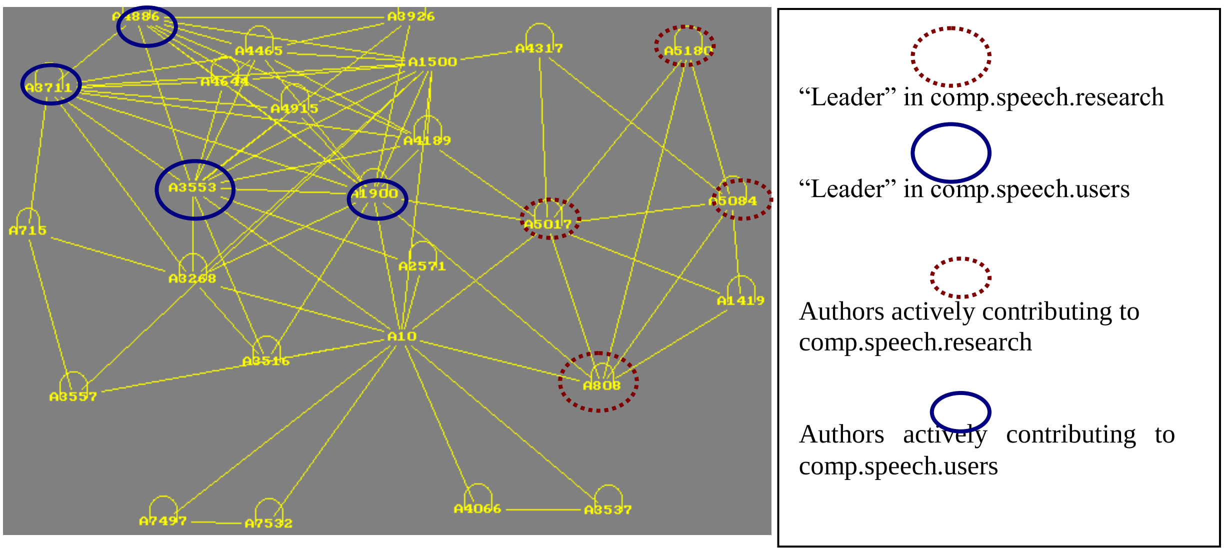
Comp.speech	Comp.speech.research	Comp.speech.users
19 cliques found.	10 cliques found.	11 cliques found.
1: A10 ANF A3553 A1900 A3516 A3268	1: A12 A7 A233 A282 ANF	1: A110 A24 A14 A62
2: A10 ANF A3553 A1900 A1500 A3268	2: A7 A197 A233 A282	2: A24 A14 A62 A3
3: A10 ANF A3553 A2571	3: A7 A233 A101	3: A24 A14 ANF
4: A10 ANF A4066 A3537	4: A7 A111 A600	4: A24 A14 A23
5: A10 ANF A1500 A3557	5: A12 A7 A600	5: A14 A3 A12
6: ANF A715 A3268	6: A381 A632 A711	6: A14 A12 A26
7: ANF A715 A3557	7: A979 A977 A982	7: A14 A257 A927
8: A7497 A10 A7532	8: A979 A977 A989	8: A110 A308 A88
9: A3553 A4886 A3711 A1900 A1500 A4189 A4465	9: A111 A600 A78	9: A110 A24 A88 A62
10: A3553 A4886 A3711 A1900 A1500 A4465 A4644	10: A12 A696 A697	10: A24 A255 A62
11: A3553 A4886 A3711 A1900 A1500 A4915		11: A1543 A206 A6
12: A3553 A4886 A1900 A4465 A3926		
13: A3553 A3711 A1900 A1500 A3268		
14: A3711 A715 A3268		
15: A1900 A4189 A5017		
16: A808 A1419 A5017 A5084		
17: A808 A5017 A5084 A5180		
18: A10 A1900 A808 A5017		
19: A4317 A5017 A5084		

[35] An extended analysis of Author "Activity" and "Prominence" is provided in a forthcoming paper from the author.

The empirical data confirm both assumptions. In the first place, the subgroups emerging in each of the smaller networks resulting from the initial newsgroup are substantially different from one another (*i.e.* contributors to speech science are normally not contributors to speech technology, and vice-versa). We may also identify "leaders" in each network, that have a special role in participating very intensively in several social circles within that network, but clearly not participating in the other newsgroup at all (A14 in comp.speech.users, and A7 in comp.speech.research, who correspond respectively to contributors A3553 and A808 in the initial comp.speech network). The majority of the other participants in each network do not participate at all in the other network. Nevertheless, and this is important for the complementary nature of both newsgroups, some members are present in both networks, allowing for the connection of these discussion fora (A111 and A12 are the same contributors as A88 and A62 respectively in comp.speech.research and comp.speech users, but coded differently for privacy reasons) .

In the following graph the clique structure of the initial newsgroup is represented. Authors exemplifying the structural evolution to specialised newsgroups are also identified.

Figure 5-2-5 - Electronic "Invisible Colleges" Emerging in comp.speech.research and comp.speech.users



In any of these networks, there is a very significant overlap of the same members in different "invisible colleges" or cliques. This is recognition of the "prominence" of those members within these electronic systems and provides evidence for the existence of a "Matthew effect" kind of process within these electronic environments.

The most significant result is the confirmation that there is a structural evolution from the initial newsgroup (comp.speech - 1992 - 1998) to the new specialised newsgroups, after May 1998. The subgroups detectable in the initial embryonic network were then extended and reinforced in the subsequent networks. Some of these subgroups are the very interactive ones occurring in the comp.speech.research network, while other distinct ones have formed in the comp.speech.users network. Given the previous evidence on the inter-sectoral nature of the collaboration within these networks - and particularly the differences in the relative weight of the university and industry sectors in the different networks - we can reconfirm the existence of a "speciation" process, now with structural evidence.

In this section we have focused on the structural characteristics of these electronic environments, and have found empirical evidence to validate the process of creation of "invisible colleges" or tightly bounded and densely connected subgroups within the diverse contributors to these networks. Further evidence has corroborated the existence of a "speciation" process in the evolution of these electronic environments, as well as reinforcing our hypothesis about a self-reinforcing mechanism in the prominence of the most active contributors to these newsgroups.

We previously identified that active researchers and technicians in the field of speech science and technology have been among the most active contributors to these newsgroups, and are particularly prominent among the select few who form the "invisible colleges" apparent from the exchanges of information within these networks. This is certainly a necessary condition for the exchange of "knowledge" within these electronic environments.

The final question concerning the potential creation, distribution and use of knowledge within these systems constitutes the topic of the last section.

5.2.7 Knowledge-Exchange within Electronic Networks

How far can these electronic environments go in the process of creating new scientific knowledge or even in the distribution and use of existing scientific knowledge? This question leads us to examine the content of the messages exchanged within these newsgroups. If we return to the discussion in the literature review section, the answer to this question is more tentative than definitive. Indeed, it is a question of degree - how intensive in scientific knowledge is a certain message or contribution? Beyond a discussion about the concept of "scientific knowledge" (something that may mean different things to different people), a practical consideration is how intensive in terms of "knowledge" or "scientific and technical information" are the information exchanges occurring in these newsgroups. Thus, the question can be reformulated as how intensive in terms of scientific and technical information are the messages exchanged. As the exact definition of "scientific information" covers a wide spectrum, from the more positivistic perspective of bibliometric and scientometric studies to the more constructivist approaches, a pragmatic approach should inform the methodology for assessing the content nature of these environments. There is, however, an important property of these electronic environments that should be stressed at the outset: these electronic fora are open and public fora for discussion, lacking any refereeing system or any similar filtering mechanism for either the content of the messages or the form of participation.

The methodology followed here was based on a content analysis of a sample of messages exchanged in the comp.speech.research newsgroup. The selection of this newsgroup was natural given its focus on research and development issues. Based on an initial inspection of a random sample of discussions, 14 different categories for classification of message content were chosen (these categories are listed in the table below summarising the results). Four different samples of discussions and respective messages were analysed, potentially representing different sectors of the population generating different "kinds" of content. These four samples of 68 discussions represent a total of 18.6% of the total of 1,462 discussions, and a sample total of 1,150 messages representing 34.8% of the total of 3,304 messages.

The first sample consists of all the most relevant or more intensive discussions. This corresponds to 68 different discussion threads and a total of 755 different messages, representing 22.9% of the total number of messages. The second sample is a random sample of 68 messages taken from the total population, excluding the top percentile of the 68 most relevant discussions. The third sample is also a random sample of 68 discussions, taken from the whole population but excluding both the most relevant as well as the discussions without reply. The last sample consists of a random sample of 68 different discussions without reply (hence, a total of 68 messages). The full text of the messages of these discussions has been read and, based on content, each has been categorised into a single content category. The categorisation process is presented in Table 5-2-XI.

Some subjective conclusions can perhaps be drawn. In general, the technical level of the discussions was high, in accordance with the kind of community participating in these fora (the scientific and technical community). There is a behavioural aspect to participation in these newsgroups; the more active contributors usually demonstrate a better knowledge of the field under discussion and are keen to give "assistance" to novices or curious observers; sometimes this "web of assistance" has elicited the contributions from other active contributors with more practical knowledge about a particular question.

Table 5-2-XI - Content Analysis of Newsgroup Discussions - Description of Procedure

Discus sID	Subject	TotalMe ssgs	Type of Discussion	Category
126	dragon nat. speak.... l&h voice xpress, ibm viavoice eval	45	Comparison between Speech recognition Software Applications: Natural Speaking; Dragon Systems; IBM Via Voice ; and L&H Voice Xpress. Discussion of Speech User interfaces and Continuous speech recognition. the Baysian principle of maximum likelyhood estimation and Word and Language Markov Models. To derive the correct meaning from context. Cepstral analysis for removing noise from data. recognizer's "certainty values" and "decoding probabilities". A typology of errors: gross acoustic mismatch (e.g. an OOV - Out Of Vocabulary word); gross linguistic mismatch (bad LM score); probably the vast majority of decoding errors : things that match OK the acoustics and linguistics. Homonyms and Homophones. The usefulness of the decoding graph scoring.	Comparing technological Applications
1361	Future Shock, an understatement!	37	Information overwhelm and overload (and Alvin Toffler future Shock analogy);Coping with the overload by being selective; Trading Knowledge for skills for doing practical things; Learning by doing (such as buying an old house and doing the repairs);perhaps technology _always_ seems overwhelming--even to its most studious practitioners - How can one _possibly_ navigate this ocean of technological change within the strictly limited dinghy of your own mind?;Practical solution to this problem;Technical articals that address a point I am wrestling with are sometimes valuable beyond price;When technology gets complicated enough it's not a tool any more;We, the human race, have created 80% of the entire body of ALL human knowledge since 1989;I hire tradesmen to do this type of stuff;I never _could_ understand those things (reading Scientific American);Ray Andraka's "Cordic" article made the subject clear for the first time;learning the Windows API. Does that count as "knowlege"?;"How may papers wre written on ...".M	General Discussion and Sociability
…	…		…	
…	…		…	

Table 5-2-XII - Summary by Category of Content Analysis Results

	Top 10%	Random All	Random but 1	Random 1 Messg
Discussing a Technical definition	4	1	4	
Discussing a Technical definition/Discussing Methodological Issues	8	2	4	2
Discussing Methodological Issues	17	16	8	8
Looking for Tools/Discussing methodologies	8	1	9	6
Looking for Tools	12	13	11	6
Comparing technological applications	5	3	3	1
Discussion of New technological applications	2	5	1	8
Looking for Experts/institutions	4	3	2	
Looking for Bibliographic References/Looking for Data/Methodologies	2	5	16	18
General Discussion and Sociability	3			
Other Information (discussion about job posts;Patents;validity info Internet)	3	4	3	3
Advertising Jobs		6	1	11
Advertising Conference		2	1	3
Solving a technical problem		7	5	2
	68	68	68	68

Some comments on the distribution of the contents of messages to different samples and categories are worth making:

- In the most intensive discussions, we usually do not have the kind of content involving passive publishing of information and lacking intensive interaction. This sample does not contain any cases in the categories "advertising jobs", "advertising conferences" or "solving a minor technical problem"; these contents are more likely to occur in the group of messages without any reply, as their frequency is substantially higher in this sample compared to the average from any other random sample.

- The discussion of theoretical issues or significant methodological issues is more likely to occur in the most intensive discussions, which are also those where the most active authors participate intensively. The categories "Discussing a Technical Definition" and "Discussing a Technical Definition/Discussing Methodological Issues" are more relevant in this sample of discussions than in any of the other samples (and are not likely to occur in a discussion without reply).
- The less intensive discussions (including both the ones without reply and other discussions with just a few reply messages) have a particular focus on "Looking for Bibliographic references/Looking for Data and Methodologies". These "search" activities are not of particular interest to the group of participants in the most debated discussions.
- The "Discussion of Methodological Issues" and "Looking for Tools" are likely to have a reasonable attention in these discussion forums, as they constitute the major proportion of the information exchanged in these electronic forums;
- Even though these are public and completely open arenas for discussion , "Looking for Experts" is a category that deserves a significant reaction and feedback, even if it occurs relatively infrequently; when such queries do occur, they generate intense discussion;

From the content analysis we do not have empirical evidence to validate the hypothesis that these systems support the *creation* of stable scientific or technical knowledge. This is certainly closely related to the absence of any "refereeing" mechanism for evaluating the content of the messages, other than the "social reputational" mechanism of being visible in a public discussion forum. Moreover, the kind of "knowledge" claims occurring in these electronic environments are more the informal type, than the formal codified form we would expect in the knowledge-creation process.

Nevertheless, we found some evidence that the structuring of these electronic communities reflects an informal "reputational" system, assigning "prestige" to the most active and prominent contributors. This is in some degree related to the discussion of more technical issues among the most active contributors. The intensity of contribution of active researchers in this field, as well as of active technicians and practitioners, is a good indication of the existence of a future "evaluation" function within these electronic discussion environments.

Along another dimension, empirical evidence was analysed corroborating the technical level of the discussions as well as the methodological nature of the content of the messages. To a large extent, the content analysis testifies that there is a good signal-to-noise ratio.

Thus, these newsgroups can be said to support the exchange of scientific and technical information and to help to create the kind of social circles in which colleagues exchange tacit knowledge to provide the solution to a problem. Most of these problems are technical by nature, but the methodological character as well as the theoretical nature of the discussions, are also manifest. The best way to describe these electronic systems would be as permanent "webs of assistance", where some reputational collegiality is present, but, nevertheless, being some way from being a formal system for scientific and technical communication. However, their role as informal channels for the dissemination and use of scientific and technical information is clearly well established.

5.2.8 Conclusions

Computer networks and electronic environments are potential facilitators of the process of scientific communication. This chapter has focused on the analysis of an especially interactive medium for scientific and technical open, public and informal communication - Newsgroups. The first part of the chapter suggests the use of indicators for assessing the overall pattern of evolution of these electronic systems, allowing the specialisation process usually found to occur in the evolution of these discussion fora and in scientific communities, to be mapped. A taxonomy of the activity and prominence of the contributors to these electronic networks, as well as the diversity of relevance of the discussions, is also suggested. The collaborative nature of these electronic-social networks is analysed in its international and inter-sectoral forms. Once again, the technological characteristics of this medium appear to facilitate remote collaboration.

In the second section, a structural analysis of the social networks resulting from the discussions provided empirical evidence to validate the hypothesis that these systems support the creation of electronic "invisible colleges", given the cliques resulting from the analysis on electronic networks for comp.speech, comp.speech.research and comp.speech.users. Moreover, the reputational mechanism of prominence seems to be well established within the community, as the leading contributors participate in several groups of interaction and increase their prominence over time (results from the network analysis of the evolution from the initial newsgroup comp.speech to the subsequent specialised newsgroups). Over time, this structural analysis reinforces the "speciation" model of evolution of these electronic communities (as different cliques or groups are involved in different specialised newsgroups).

Finally, the content analysis of a significant number of discussions has suggested that these electronic networks are being used as "webs of assistance" and informal channels for the exchange of already existing scientific and technical knowledge of a more tacit nature. The reputational system for social interaction within these systems seems to ensure a reasonable signal-to-noise ratio in the information and knowledge contents distributed and used. Particularly important is the discussion about methodological issues. However, the lack of any "refereeing system" equivalent to the role of "peer review" in the formal communication system of science, leaves these electronic systems in the realm of the informal communication system of scientific communities.

A further line of investigation in this area is related to the comparison of these results with several other scientific and technical communities and more focused analysis of the content of information exchanged within these electronic environments. Indeed, further research is needed on the classification of the information accumulating within these electronic networks as well as the typologies that should be used to categorise content. These are important topics for further investigation.

The following section specifically addresses this problem of examining Internet networks as infrastructures supporting the dissemination and distribution of information. Moreover, the links among entities inter-changing information within these networks are examined to see if patterns of non-electronic interaction and collaboration are reproduced.

5.3 Mapping Research Networks in Internet Space. The Discovery of Digital Knowledge Bases

5.3.1 Introduction

This is the third and concluding section on the direct analysis of changing patterns of scientific communication and research collaboration due to the availability of advanced information and communication infrastructures. In section 5.1, we analysed, from the viewpoint of individual researchers, how a combination of several ICTs is being used for different stages of the research process. In addition, we studied how this technological infrastructure is likely to enhance research collaboration, namely remote collaboration and collaboration within inter-disciplinary areas of research. Secondly, in section 5.2, we pursued a detailed investigation of the collaboration structure of particularly interactive electronic discussion fora. The longitudinal analysis of Newsgroups on speech science and technology allows us to conclude that these informal systems of communication are indeed extending "invisible colleges" into the electronic infrastructure of science. They support inter-sectoral and international collaboration as well as the usual "division of labour" in the practice of research.

We turn in this section to analysis of the electronic structure of research networks and the dissemination of information in electronic environments. The empirical data and discussion are based on analysis of the ELSnet network. This section addresses the problem of mapping the structural linkages of research networks on the Internet for the purpose of identifying knowledge bases on electronic networks. Traditional (non-electronic) research networks are likely to have a digital representation (a web presence), whose boundaries and characteristics require closer investigation. Of special concern here is to identify particular sub-sets of these digital networks whose properties are related to non-digital collaboration structures.

In the first place, we are interested in testing the hypothesis that the patterns of connectivity of these electronic networks are structurally similar to the collaboration patterns identified in section 4.2. There, we discussed the research project collaboration structures that have evolved following a decade of European Commission funding of this research field, and the inter-personal collaboration structures revealed by a survey of ELSnet researchers. We find that these electronic connectivity patterns allow a restricted set of "best connected" research institutions and the linkages to important external entities and electronic resources to be identified.

Secondly, when analysing more deeply the inner structure of these "web communities", we can identify different degrees of centrality and prestige even among the "core" of very well electronically connected research institutions. At this point, we need to try to obtain a better understanding of how many other institutions are directly linking to each institution as well as "who is" directly connected to each of the 141 ELSnet research groups. As a complementary measure of the centrality of each institution in the whole network, we also mapped the electronic "ego-networks" (direct and indirect links starting in a specific institution and spreading into the wider Internet space) for some of the most central research institutions. Again, these centrality results corroborate the collaboration patterns of the actual non-electronic network.

Thirdly, having characterised the connectivity pattern of these electronic networks, as well as the heterogeneity in centrality of different institutions within the larger network, we explore the possibility of identifying particularly intensive "knowledge bases" on electronic networks. By mapping the ego-networks of a reasonable number of these better connected and highly central institutions, we might be able to identify "overlapping regions" of very intensive knowledge resources topically related to this field of research. A conceptual model is represented in this final section.

The analysis is organised in the following way. In section 5.3.2 we situate the topic of the "link structure" of the Internet and the discovery of knowledge bases and provide an overview of the literature in this area. Sections 5.3.3 and 5.3.4 describe, respectively, the research issues and hypotheses and the methods used for conducting the overall analysis. The following three sections correspond to the detailed analysis and discussion of the three topics outlined above in the present introduction. Finally, we conclude this section by summarising the most significant results and limitations of this empirical investigation and point to avenues for further research.

5.3.2 The Link Structure of the Internet and Digital Knowledge Bases

In this investigation we use techniques of link-structure analysis for determining the electronic connectivity of research institutions and the centrality of some institutions in the overall network and devising new models for discovering knowledge resources on these sparse and disconnected networks. This is closely linked to both search-engine technology applications and to recent studies exploring the topology of the Internet to identify web communities. We will examine these two topics before moving on to a more sociological exploration of this problem.

Search-engine technology is mainly concerned with providing a solution to two related functionalities: *precision/relevance* and *recall*. The results of a search on the Internet should be as precise as possible in order to be effective. From the huge, and perhaps impossible to determine, collection of information available, it should, in a few seconds, produce more precise results on the topic being searched. This goal is sometimes at odds with the number of items of information that it can give as an output to the user (the recall capacity of the search engine). We would expect that the *more* results returned, the better. In other words, search engine technology should provide both more items and very precise items of information from a search.

Considering the huge size of the web and its dynamic nature this is a very complex technological undertaking. Most search engines index documents in a database recovered from systematic crawls of the web. The crawls explore the link structure and hypermedia nature of the Internet in order to jump from one document to related documents. Complementary content analysis provides a classification of the retrieved documents. However refined search-engine technology has come to be, estimates suggest that 12 leading search-engines taken together index less than 50% of the whole internet system (Lawrence and Giles, 1999).

As noted by Aquino and Mitchell (2001), search-engine technology has used three different types of algorithms to build these indexed databases: the Naïve Bayes model, which focuses on topic-word frequencies; "maximum-entropy" algorithms, which focus on word combinations and how frequently they are associated; and, perhaps the most promising approach, the "co-training" model , which studies the information on a web page as well as the linked pages, building an association of correlations. In fact, these strategies of combining the content of Internet resources with the link structure of those resources are better suited for "entity extraction", or the building of databases from collections of specific entities.

This brings us closer to the notion of "web communities" and the self-organisation of the Internet. If we consider a web community as a set of web pages that link (in either direction) to more web pages within the community than outside of the community (Flake *et al.* 2000), it is important to be able to identify such subsets in the vast electronic network.

Several studies have empirically identified such communities based on a combination of algorithms. One of the common ones is the HITS algorithm (Kleinberg, 1997), which explores the link structure of the Internet, starting from a set of seed URLs and determining the *Hubs* and *Authorities* resources on Internet web space. Hubs are Internet resources that link to many authoritative pages in the topic, while Authorities are Internet resources that are linked by many Hubs. There is a self-recurring mechanism in the idenfication of Hubs and Authorities. This problem has been overcome by refinements to the initial algorithm. The method is based on partitioning the initial graph into well defined components. The determination of Authoritaties pages on a particular topic is quite useful when ranking the results of search queries.

The PageRank algorithm, implemented in the search-engine GOOGLE, uses a very refined version of the initial HITS algorithm (Brin and Page, 1998; Huang, 2001). This is important to note as we use Google technology for determining the collection of Related pages in our initial research network.

From a technical point of view, theoretical and empirical research have demonstrated the effectiveness of using the link structure of the network to identify subsets of the wider network that can be categorised as "web communities". However, analysis of the inner structure of these "web communities" is still challenging and probably the solution is not particularly technical but more sociological in nature. In fact, we should try to understand the organisation of these electronic communities by comparison with other structural characteristics, such as the similarity of the information under exchange within these communities, or the collaboration structure of the institutions participating in these communities.

Sociological explanations of on-line virtual communities have particularly focused on user studies of the behaviour of those communities when participating in these electronic environments (Wellman, 1996; Garton *et al.*, 1999; Koku *et al.*, 2000). Much less research has been done on the automatic examination of electronic collaboration structures, as given by Internet resources and archives, at any level higher than the individual. Previous research has focused essentially on analysing particular events (such as electronic conferences, newsgroups participation), but not large-scale electronic infrastructures.

There does not appear to be any study that seeks to explain the sociological factors underlying the electronic connectivity of a set of institutions. The particular focus of this research is to start the analysis with a well-bounded and restricted set of research institutions (relevant for participation within the same research network). This focused search is likely to impose an *a priori* restriction on the total electronic webspace, which might yield very good results in terms of topical focus and discovery of knowledge resources. In addition, we explore the idea that the electronic webspace of the community should, in a particular way, be related to the sociological structure of the community. Connectivity patterns and the centrality of some institutions within the electronic network should also reflect this. These are all research questions that are apparently not covered by the existing literature and which involve important theoretical hypotheses.

5.3.3 Hypotheses to be investigated

One crucial set of research questions is whether the electronic patterns of connectivity of a certain set of research institutions reflect the structures of research collaboration of the physical (non-electronic) research network. If so, what kind of structure is revealed by the electronic network? Can we identify the best connected institutions and their characteristics?

Related to these questions is the possibility of discovering very intensive knowledge resources on electronic networks by mapping the topology of densely connected and central institutions in Internet space - we will refer to these as "digital knowledge-bases".

The analysis is conducted with a view to testing the following theoretical hypotheses:

a) That the pattern of electronic connectivity for the set of research institutions is structurally similar to the already analysed collaboration structure within the ELSnet network.

This should allow a core of better connected institutions to be identified within electronic space and within this group to identify certain properties that explain this connectivity pattern. Moreover, the network mapping of electronic relationships should allow identification of institutions that collaborate more intensively with each other in traditional settings.

b) That even among the better connected "core" group of institutions, we are able to identify different degrees of centrality and prestige, as measured by the number and type of institutions connected to them (*indegree centrality)* and their activity in electronic space - the number and type of links originating from them (*outdegree centrality*).

This allows identification of some heterogeneity within the network characterising the inner structure of the electronic map and leads on to the final hypothesis related to the potential identification of knowledge bases in electronic networks:

c) That the mapping of ego-centered networks (the links originating in a certain institution and extending into a specified depth) of the most central and best connected institutions enables identification of digital knowledge bases.

Assuming that hypotheses a) and b) are validated, i.e. the electronic connectivity resembles the collaboration patterns of the institutions involved, and that it is possible to distinguish the most central and more active institutions, then we might explore the idea that detailed examination of their electronic links allow important knowledge resources to be found on the Internet. As this final hypothesis is exploratory, we discuss here a conceptual model for the identification of these "digital knowledge bases".

5.3.4 Methods for Mapping Research Networks on the Internet

The empirical analysis will focus on the examination of the electronic connectivity of the member institutions of the ELSnet network. This network involves 141 research groups/institutions in 27 countries across Europe. Most of these institutions participated in the Research Programmes and Projects funded by the European Commission over the period 1990 - 2002 (see the detailed structural analysis in Chapter 4.2). These collaboration structures, identified previously by network analysis, are used as a proxy for the collaboration structure within the whole of the ELSnet network. ELSnet network researchers comprised the population used for the survey on collaboration and use of ICTs

- the topic of section 5.1. The inter-personal collaboration structures, as given by the institutions with whom researchers collaborate more frequently, is used as a proxy for the collaboration structures within the network.

The analysis of electronic connectivity of the ELSnet institutions is based on data collected from the search-engine GOOGLE (available at http://www.google.com). This search-engine technology was chosen for three reasons: 1) this search-engine repository is considered to be one of the largest indexed databases (ShowdownSearch, 2002); 2) the technology and algorithms underlying the search process are detailed in the literature and are particularly suitable for the analysis of web communities; 3) there are two specific functionalities that are exclusive to this search engine - advanced search of all the pages that link to a specific page - *Link:* feature; and advanced search of a set of approximately 30 pages that are strongly related to a specific page - *Relate:* feature .

The methodology, based on a combination of webmetric techniques and network analysis, consisted of the following procedures:

- Collection and analysis of all the institutions that are strongly related a particular research institution (as given by the search engine technique described above);
- Detailed network analysis of the connectivity of the resultant network of inter-related research institutions (in webspace).
- Comparison of the electronic connectivity network with the patterns of collaboration revealed by the physical (non-electronic) ELSnet network.
- Collection and analysis for each research institution of all the institutions that are linked directly to it (as given by the link: pages feature of Google as described above).
- Characterisation of the centrality and prestige of the research institutions as given by their position in the electronic network and total number of links pointing to them and type of institutions linking to them.
- Collection of data for the ego-centered electronic networks of a sample of very well connected and central research institutions.[36]
- Production of a Network map of these ego-networks to facilitate a visual inspection of their tree-like structure and to derive the conceptual model for the final section.

[36] A customised Visual Basic software application was developed by the author in order to *crawl* along the Internet linkages related with a certain institution's URL, to a determined depth of analysis. A more detailed description of this instrument is available at the following URL, as of 5^{th} of June 2002 (http://www.sussex.ac.uk/Users/prpb7/DigitalTools/Galilei/galilei.html) and can also be obtained direct from the author.

The discussion of the results follows in the next three sections, dealing respectively with the patterns of electronic connectivity, the centrality and prestige of research institutions in electronic space and the conceptual model for the discovery of digital knowledge bases.

5.3.5 Patterns of Electronic Connectivity of Research Networks

We start by outlining certain methodological limitations. From the total set of 141 research institutions participating in the ELSnet network, we were able to analyse the patterns of electronic connectivity of 108. The remaining 33 institutions were not eligible for several reasons: 1) 13 institutions did have a URL identification, but the URL referred to the global institution and not to the specific research centre or group focusing on speech and language research or development. This generally applied to large R&D companies (e.g. Nokia research, IBM Deutschland, DaimlerChrysler AG, Philips Electronics, BT Adastral Park), but also to some university faculties (e.g. University of Lisbon - Department of Informatics and Università degli Studi di Pisa); 2) 12 institutions did not apparently have a URL, at least as given by the list of ELSnet members (in some of these cases, they probably do have a URL but it is too generic to represent the specific research group; and 3) 8 of the institutions did not have any relationship result, as given by the Google *relation:* functionality.

For the research institutions analysed, we were able to determine a list of related entities (URLs). This list of related entities is restricted to a maximum of 30 URL references. These references point to strongly related internet resources (web pages). The measure of relatedness is given by a combination of structural relationships – a set of linkages characterising two related institutions as well as topic/content similarity. For the present purposes, we consider this indicator of relatedness as a measure of electronic connectivity between institutions. It should be stressed that the list of related institutions is not restricted to members of the ELSnet network. Even though our main purpose is the identification of the connectedness between members of the network, most of the "links" of relatedness were to outside entities and URL references. In fact, some highly relevant external links to "authoritative" URL references are one important characteristic of the best connected institutions (for more about this, see the final part of this section).

From the analysis of this connectivity dataset, we were able to identify a "core" set of 24 institutions with a denser relationship with other members of the ELSnet network. A combination of three important properties distinguishes this reduced set of institutions.

First, as noted above, all of them show a pattern of more intense electronic relationships with other members of the ELSnet research network. Secondly, they also have electronic relationships with institutions that are not part of the ELSnet network but are working in the same research and development field. Thirdly, as noted above, they have electronic relationships with important resources, which are external to the network, but which are considered hubs and authorities (in the sense defined in the literature review - see section 5.3.2) - i.e. internet resources pointed to by many other URL references and linking to other authoritative resources. Table 5-3-I shows the list of the best electronically connected institutions.

Table 5-3-I - The 24 Research Institutions more Densely Connected, as Given by Electronic Connectivity

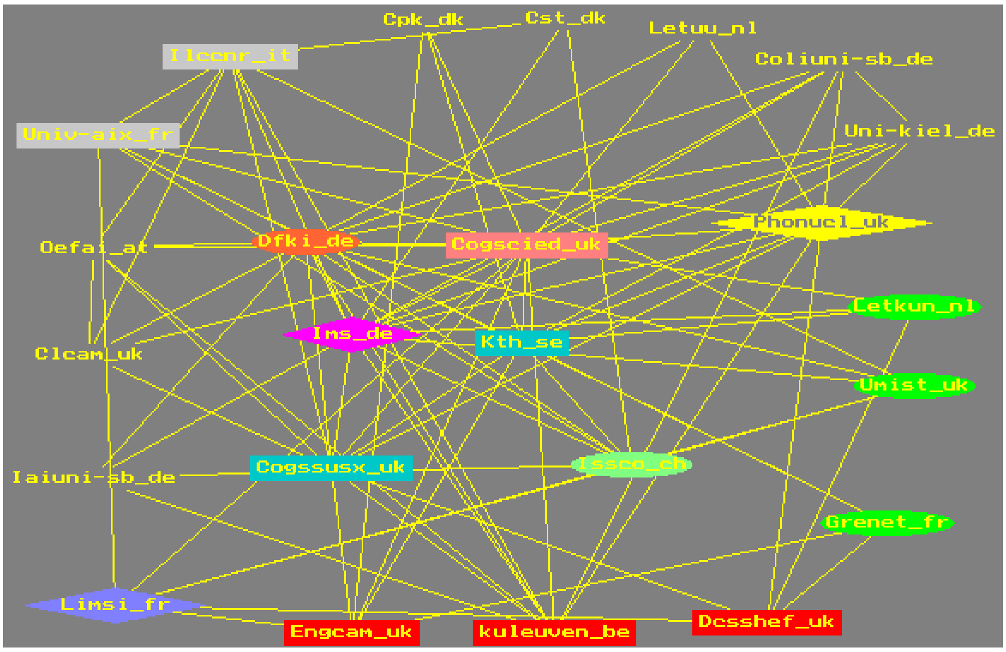
Best Electronically Connected Institutions	
Research Institutions	***URL***
Austrian Research Institute for Artificial Intelligence (OFAI)	http://www.ai.univie.ac.at/oefai/nlu/
Katholieke Universiteit Leuven - Centre for Computational Linguistics (CCL)	http://www.ccl.kuleuven.ac.be/
Aalborg University - Institute of Electronic Systems	http://www.cpk.auc.dk/
Center for Sprogteknologi	http://www.cst.ku.dk
Inst. National Polytechnique de Grenoble - Institut de la Communication Parlée	http://www.icp.grenet.fr/
LIMSI/CNRS - Human-Machine Communication Department	http://www.limsi.fr/
Université de Provence - Laboratoire Parole et Langage	http://www.lpl.univ-aix.fr/
Christian-Albrechts University, Kiel - Institute of Phonetics and Digital Speech Processing	http://www.ipds.uni-kiel.de
German Research Center for Artificial Intelligence (DFKI) - Language Technology Lab	http://www.dfki.de/lt
Institut für Angewandte Informationsforschung	http://www.iai.uni-sb.de
Universitaet des Saarlandes Department of Computer Linguistics	http://www.coli.uni-sb.de
Universität Stuttgart-IMS - Institut für Maschinelle Sprachverarbeitung	http://www.ims.uni-stuttgart.de
Consiglio Nazionale delle Ricerche - Istituto di Linguistica Computazionale	http://www.ilc.pi.cnr.it
KTH (Royal Institute of Technology) - Department of Speech, Music and Hearing (TMH)	http://www.speech.kth.se/
University of Geneva - ISSCO/ETI	http://www.issco.unige.ch/
University of Nijmegen - Department of Language and Speech	http://lands.let.kun.nl
Utrecht University - Utrecht Institute of Linguistics OTS	http://www-uilots.let.uu.nl
UMIST - Computational Linguistics (old URL)	http://www.ccl.umist.ac.uk/
University College London - Department of Phonetics and Linguistics	http://www.phon.ucl.ac.uk
University of Cambridge - Computer Laboratory	http://www.cl.cam.ac.uk/
University of Cambridge - Speech, Vision and Robotics Group	http://svr-www.eng.cam.ac.uk/
University of Edinburgh - Informatics	http://www.cogsci.ed.ac.uk/
University of Sheffield - ILASH, Computer Science	http://www.dcs.shef.ac.uk/research/ilash
University of Sussex - School of Cognitive and Computing Sciences	http://www.cogs.susx.ac.uk/lab/nlp/index.html

Comparison of this list of institutions with the list of research institutions appearing in the collaboration structures identified in section 4.2 (based on structural patterns of research collaboration as given by research funding of this field by the European Commission, as well as by inter-personal collaboration maps resulting from the researchers survey) reveals a strong similarity.

However, one characteristic this list is that all are research institutions from university departments or government laboratories. Thus, to a very large degree, we were able to identify from the electronic connectivity patterns, the best connected "research" institutions, but not the R&D companies involved in this research area. This is probably due to the methodological constraint referred to initially, but also because of the "higher level of secretiveness" of these companies in relation to making their research information public. This might represent a problem in the form of an electronic under-representation of this group in terms of any structural anaysis of large-scale systems such as the Internet.

In order to map the connectivity pattern of this reduced network, we carried out a network analysis of electronic connectivity. The results are depicted in Figure 5-3-1.

Figure 5-3-1 - Visual Representation of Electronic Connectivity Network
(24 core Members of ELSnet Network)



From the above network map, it is clear that all the institutions reveal a reasonable degree of connectedness with other members of the network. However, it is also possible to visualise the heterogeneity within this reduced set, in terms of degree of connectedness and centrality. For example, institutions such as the Universität Stuttgart-IMS - Institut für Maschinelle Sprachverarbeitung (*Ims_de*), the German Research Centre for Artificial Intelligence (DFKI) - Language Technology Lab (*Dfki_de*), the KTH (Royal Institute of Technology) - Department of Speech, Music and Hearing (*Kth_se*) and the University of Edinburgh - Informatics (*Cogscied_uk*)[37] occupy more central positions within the network.

[37] At first, the connectivity result for the University of Edinburgh was somewhat surprising. However, closer inspection revealed that this node contains a cluster of three very prominent research groups in this

They provide a bridge between subgroups interacting in the network. This higher degree of centrality is more fully discussed in section 5.3.6 below.

A final comment on the results of the pattern of electronic connectivity of these best-connected institutions concerns their linkages with external resources. A vast majority (all but two) of the 24 best connected institutions were related to external Internet resources. The striking result is that, even within this reduced set, the more densely connected institutions, such as the ones referred to above, pointed to the same "authoritative" institutions, giving credibility to the *reinforcing mechanism* statement that the more authoritative resources tend to be reinforced by linkages with other authoritative resources. The better connected , the more likely an institution is to be structurally related to other very well connected Internet resources. In other words, a form of electronic "Matthew Effect" is likely to occur within these electronic networks. Table 5-3-II summarises the most relevant "external" linkages.

Table 5-3-II - Structurally significant external connections (Hubs, Authorities and Other Internet Resources)

Hubs, Authorities and Internet Resources	
Hubs	
ELSnet Network	http://www.elsnet.org
ELRA/ELDA (European Language Resources Association)	www.icp.inpg.fr/ELRA/home.html
HLT (European Commission Human Language Technologies)	www.hltcentral.org/page-56.shtml (EUROMAP)
ISCA (International Speech Communication Association)	www.isca-speech.org/
Linguistic Data Consortium	www.ldc.upenn.edu/
ETSI (European Telecommunication Standards Institute)	www.etsi.fr/
International Phonetics Association	www2.arts.gla.ac.uk/IPA/ipa.html
American Speech Language and Hearing Association	www.asha.org/
Association for Computation Linguistics	www.aclweb.org/
Authorities	
Stanford Linguistics Research	www-linguistics.stanford.edu/
SRI Natural Language Program Research	www.ai.sri.com/natural-language/
Ctr.Language & Speech Processing at John Hopkins University	www.clsp.jhu.edu/index.shtml
Perceptual Science Lab	mambo.ucsc.edu/
Speech at Carnegie Mellon University	www.speech.cs.cmu.edu/speech/
Microsoft Natural Language Processing Home	research.microsoft.com/nlp/
University of Rochester Department of Linguistics	www.ling.rochester.edu/
MIT Artificial Intelligence Lab	www.ai.mit.edu/
Speech Recognition Group at Rutgers University	www.caip.rutgers.edu/ARPA-SLT/
Resources	
PRINT SERVER Computational Linguistics	xxx.lanl.gov/cmp-lg/
VIRTUAL Library of Linguistics	www.emich.edu/~linguist/www-vl.html
Comp Speech Newsgroup Website	www.speech.cs.cmu.edu/comp.speech/
Subgroup of the Association for Computation Linguistics	www.sigdial.org/
Natural Language Translation Specialist Group	www.bcs.org.uk/siggroup/nalatran/nalasupp.htm
The Linguist List - Mailling List	www.emich.edu/~linguist/
Stanford Grammar Resources	hpsg.stanford.edu/

It is significant that we were able not only to identify the best-connected institutions by analysing the electronic connectivity of the initial research network, but also that these best-connected entities are electronically connected to important topic-related external resources in the larger Internet network. This is particularly significant when discussing the

field: the Centre for Speech Technology Research, the Linguistics Department and the Cognitive Sciences research group.

potential discovery of digital knowledge bases by mapping the electronic connectivity "ego-maps" of the most relevant institutions. Moreover, it is a good indication of heterogeneity in the kind of knowledge resources we are able to identify when starting the analysis with a close-knit, bounded research network and then expanding the electronic network to include "external" highly relevant Internet resources.

In order to gain a better understanding of the inner structure of these electronic networks, we will discuss in the next section the measures of centrality and prestige of the institutions under analysis. In this section we will be interested in differentiating the institutions in terms of activity and prestige, even within the more reduced core set of "best connected" institutions.

5.3.6 Centrality and Prestige within Electronic Networks

Measuring the Prestige of Research Institutions on Internet WebSpace

A common indicator of the "prestige" of a certain entity within a network is given by the total number of links that entity/institution receives from other institutions within the network. Table 5-3-III shows the results for the in-degree centrality of the already selected best-connected institutions within the ELSnet network. It should be noted that these include incoming links from institutions outside the network, as well as "inside" links from within the same institution (an indicator of "inbreeding" or self-citation) and, as such, this indicator should be taken merely as a proxy for the importance of the institutions in electronic space in general and not as a measure of the network in particular.

Table 5-3-III - Indegree-Centrality of ELSnet Best Connected Institutions

Best Electronically Connected ELSnet Institutions - Indegree Links	
Research Institutions	***Number of Direct InLinks***
Austrian Research Institute for Artificial Intelligence (OFAI)	80
Katholieke Universiteit Leuven - Centre for Computational Linguistics (CCL)	258
Aalborg University - Institute of Electronic Systems	486
Center for Sprogteknologi	126
Inst. National Polytechnique de Grenoble - Institut de la Communication Parlée	344
LIMSI/CNRS - Human-Machine Communication Department	1660
Université de Provence - Laboratoire Parole et Langage	866
Christian-Albrechts University, Kiel - Institute of Phonetics and Digital Speech Processing	70
German Research Center for Artificial Intelligence (DFKI) - Language Technology Lab	498
Institut für Angewandte Informationsforschung	144
Universitaet des Saarlandes Department of Computer Linguistics	1030
Universität Stuttgart-IMS - Institut für Maschinelle Sprachverarbeitung	1090
Consiglio Nazionale delle Ricerche - Istituto di Linguistica Computazionale	306
KTH (Royal Institute of Technology) - Department of Speech, Music and Hearing (TMH)	1050
University of Geneva - ISSCO/ETI	454
University of Nijmegen - Department of Language and Speech	86
Utrecht University - Utrecht Institute of Linguistics OTS	534
UMIST - Computational Linguistics (old URL)	234
University College London - Department of Phonetics and Linguistics	974
University of Cambridge - Computer Laboratory	2180
University of Cambridge - Speech, Vision and Robotics Group	1060
University of Edinburgh - Informatics	542
University of Sheffield - ILASH, Computer Science	334
University of Sussex - School of Cognitive and Computing Sciences	160

The results corroborate the "centrality" and "prestige" of the best connected institutions, and this applies to those with a higher degree of connectivity as given by the analysis in the preceding section (such as DFKI_dk, IMS_de, KTH_se, and .COGSCIED_uk).

It is also noteworthy that some institutions, even if not in the restricted group of the very best-connected, are quite "visible" from outside the network. This is the case of both research groups from the University of Cambridge, (the Computer Laboratory and the Speech, Vision and Robotics Group) and the Universitaet des Saarlandes Department of Computer Linguistics.

Moreover, there is a certain number of institutions that, though they do not belong to the group of electronically best-connected within the ELSnet network, do have a very high degree of "visibility" in relation to the external network. This is important for they might be considered as "authoritative" resources from outside the boundaries of ELSnet. This is the case for the Université Paul Sabatier (Toulouse III) - Institut de Recherche en

Informatique de Toulouse (IRIT), Xerox Research Centre Europe - Grenoble Laboratory, Langenscheidt KG - Electronic Media from Germany, Universität des Saarlandes CS-AI - AI Lab at the Dept.of Computer Science, and the Institute for Language & Speech Processing (ILSP) in Greece. This result is also important for the discussion of modelling the discovery of digital knowledge bases, as this should indicate that not only should the core group of densely connected institutions be extensively mapped, but also some of the less well connected ones within the network that are "authoritative" in terms of external links.

These results led to a more detailed analysis of "Who is linking directly to a given institution's electronic site?". Analysis of a sample of the electronically best connected institutions as well as some of the ones revealing more direct inlinks provided some interesting results, summarised in Table 5-3-IV.

Table 5-3-IV - Who is Directly Linking to One's Institution

Direct InLinks to sample of best connected and most "prestigious"

	Percentage of Direct InLinks (by type)				
Research Institutions	***"Inbreeding"***	***Best Connected***	***Other ELSnet***	***Hubs /Authorit***	***Other***
LIMSI/CNRS - Human-Machine Communication Department	58,6	4,2	7,4	2,4	27,3
German Research Center for Artificial Intelligence (DFKI) - Language Technology Lab	55,0	11,3	4,6	6,6	22,5
Universitaet des Saarlandes Department of Computer Linguistics	57,8	10,3	5,3	3,2	23,4
Universität Stuttgart-IMS - Institut für Maschinelle Sprachverarbeitung	26,5	11,9	8,8	7,5	45,2
KTH (Royal Institute of Technology) - Department of Speech, Music and Hearing (TMH)	48,5	4,2	3,3	3,9	40,1
University of Cambridge - Computer Laboratory	50,3	2,5	2,0	2,4	42,9
University of Edinburgh - Informatics	26,4	9,4	5,1	6,3	52,8
University College London - Department of Phonetics and Linguistics	56,0	3,1	4,2	3,5	33,2

One clear conclusion from these results is the high percentage of self-citations or direct links coming from inside the research institution's own homepages. This is an indicator of electronic "inbreeding". Between 50% and 60% of the total number of links connecting certain institutions come from inside that institution. This is valid for all but two cases among the most very well connected institutions (the Universitat Stuttgart - IMS and the University of Edinburgh - Informatics). It comes as no surprise that the best connected institutions should also be less "inward looking" institutions. But it is very satisfying to confirm this conceptual assumption.

It is also noteworthy that a significant proportion of the self-citation links refer to the institutional internet resources of Staff information and Project information. This is extremely relevant for the discovery of electronic knowledge bases (to be discussed in the next section).

As we have focused our analysis on the best electronically connected and most "central" institutions, it comes as no surprise that the percentage of direct inlinks coming from other ELSnet best-connected institutions is significant and tends to be higher for the very best connected ones. Another interesting finding that reinforces the results from the preceding sections is that the very best connected institutions tend to cite each other directly. This evidence reinforces the assumption of the existence of "local proximity in electronic space" among the best connected institutions. They form a dense subgroup of highly connected institutions, revealing a kind of "prestigious accummulation process". This process is again reinforced by the relatively high percentage of external entities (hubs, authorities and Internet resources in the field) that link directly to those institutions. It is apparent a self-reinforcing mechanism that "success breeds centrality and centrality breeds success or prestige".

It is also important to stress the direct connectivity from other member institutions of the ELSnet, which accords with the idea that this core set of institutions function as "internal authorities" to the network. Thus, here again, the assumption is confirmed that "the winners structure the network", or at least apparently have a strong influence on the structure of interactions within the network.

A final comment on the results is that there are a significant number of links coming from other miscellaneous institutions (around 30% - 45%), some of which are related to the field of research and development but some of which are not. This connectivity is important for ensuring some heterogeneity in the knowledge discovery process (see section 5.3.7).

Measuring the Electronic Activity of Research Institutions
Out-degree-Centrality

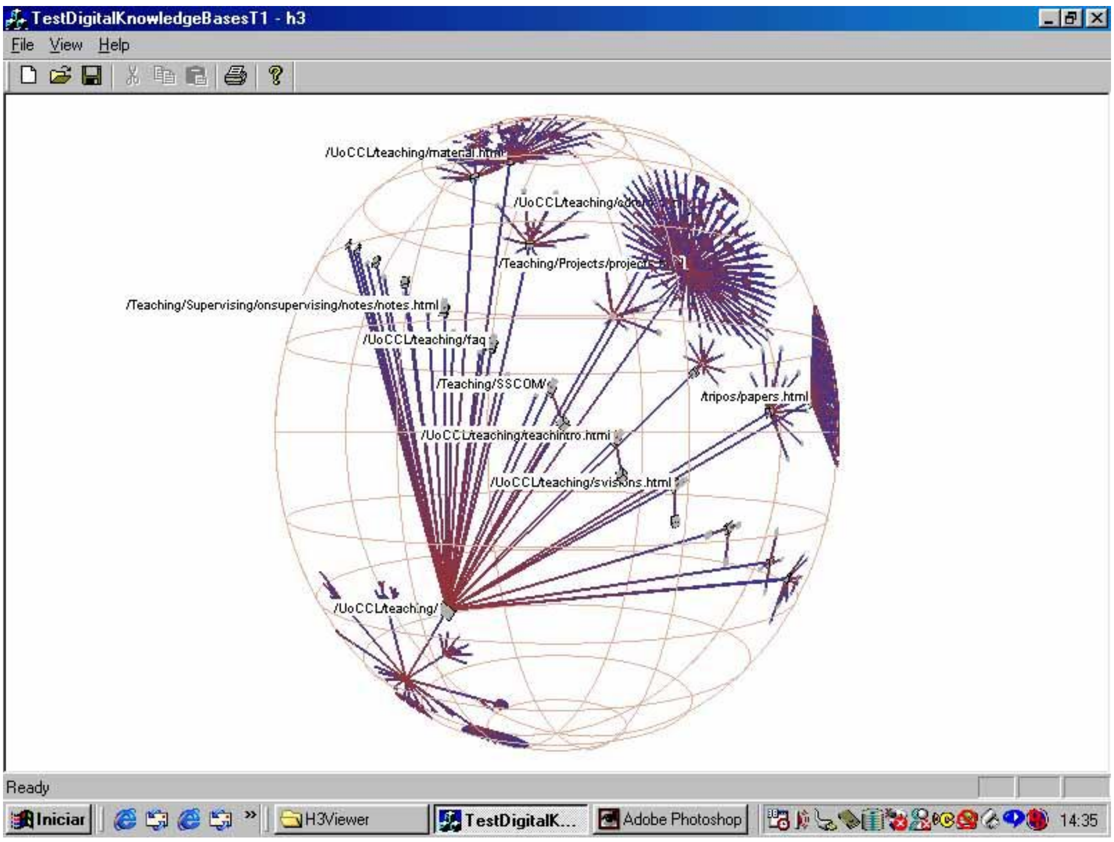
Another complementary measure of the centrality of institutions within a network is given by the out-degree centrality, or number of links originating in a certain institution. Given the hyperlinked nature of the Internet this indicator is not measurable unless the *depth* of the outgoing links is restricted to a certain *n* degrees of connectivity. By doing this we can compute and visualise the ego-centred network of any institution, to a depth *n* of analysis.

With some similarity to the previous analysis of direct incoming links to a certain institution, the analysis of outgoing links also reveals important characteristics of the connectivity of institutions. From the analysis of the electronic ego-networks of a restricted sample of the most well connected institutions, the results confirm interesting expectations.

- The proportion of outgoing links to information resources within the same institution is quite significant (at least 60-70% of outgoing links at depth 4). So, here again, we find a significant percentage of "inbreeding" and, again, this is strongly correlated with information on staff and projects, as well as internal "knowledge-bases" (e.g. papers, reports and so on).
- We can clearly identify strong connectivity with those institutions with which an institution collaborates more intensively (particularly in research projects and in the inter-exchange of reseach staff, such as for PhD or postdoctoral work).
- We can identify external linkages to important authoritative resources.

Figure 5-3-2 shows the results for the Computer Laboratory at the University of Cambridge.

Figure 5-3-2 - Mapping Individual Research Institutions Electronic Ego-networks



The ego-network electronic map reveals a concentration of some important information resources within the institution. Those "clusters" of information are related to researchers' published papers and on-line teaching resources as well as projects undertaken by the research groups.

At a more distant "layer" of connectivity (*n* greater than 5) we can also map links to external authoritative resources and to other members of the research network.

5.3.7 A Conceptual Model for the Discovery of Digital Knowledge-Bases

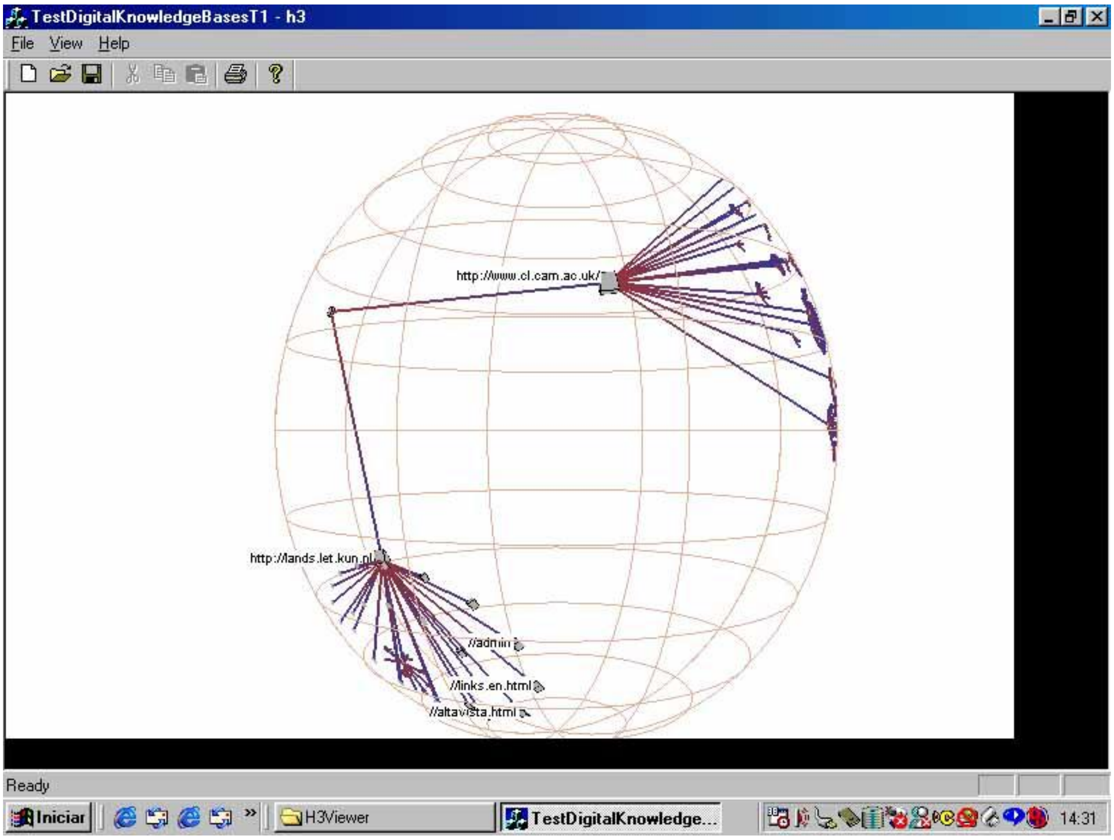
This section is exploratory in nature as the empirical evidence based on the reduced sample of five electronic ego-networks analysed is limited. Nevertheless, the results from the preceding two sections are quite revealing as to the importance of analysing the topology (link-structure) of research networks to gain a better understanding of the structure of collaboration between those institutions and the information resources made available in electronic space.

From the outset we should stress that this analysis is generally applicable only to those types of institutions that openly disclose their information and "knowledge-resources" in a publicly accessible electronic form. In a previous section (see patterns of electronic connectivity above), the empirical data reveal that private companies do not usually publish information on-line. Thus, we should remember that these results are generally only applicable to research institutions from the university and government sectors.

Among this subset of research institution members of a known network of collaboration in the field of research, we are likely to find reasonable heterogeneity in terms of patterns of connectivity, centrality and prestige within the network. This has been discussed in the preceding two sections. The empirical evidence validates the existence of subgroups within an *inner structure* of these electronic networks that are more strongly connected with each other, but that also maintain a reasonable number of links with members of the non-electronic network and link to external "authoritative" resources. Detailed examination of some of these very well connected and central institutions reveals disclosure in electronic webspace of important knowledge resources, such as publications that are available in electronic form, research project documentation and information about research staff.

The following figure shows the mapping of two electronic ego-centred networks, corresponding respectively to two institutions that are part of the best-connected entities, but, in this case, are not directly connected to one another, but link to other zones of the wider electronic network.

Figure 5-3-3 - Mapping Overlapping Components of these Ego-centered Research Networks



By mapping extensively in electronic webspace these ego-centred networks, we are able to detect “overlapping regions” where information resources from these institutions are linked together (e.g. the Computer Laboratory with the Cognitive Sciences department of the University of Sussex and the Computational Linguistics department of the University of Sheffield; and the University of Nijmegen with the University of Stuttgart - IMS and also the University of Sheffield).

A more general procedure for detecting very intensive electronic zones of knowledge-resources, related to a specific field of research, can be described in the following way:

- Start with a well defined set of institutions that, through their collaborations, reveal a structural pattern of connectivity - this is usually the case with research networks.
- Collect information about the electronic centrality of all those institutions, particularly the number and type of direct in-links;

- Map the electronic ego-centred networks of all those institutions, to a certain "layer" of connectivity, at least 10 links in depth;
- Selectively, map only those electronic linkage structures for the institutions revealed to be best connected as well as more "central" and "prestigious";
- Detect the "overlapping regions" of connectivity between the whole set of ego-networks, as well as the most relevant external linkages linking to/from members of the network (good candidates for external "authorities" and "hubs").
- Map the ego-centred network of those external "Hubs" and "Authorities"

This enables identification of digital knowledge bases or, in other words, zones of the electronic networks that are very intensive in information resources, and are simultaneouly linked to important members of the scientific research network.

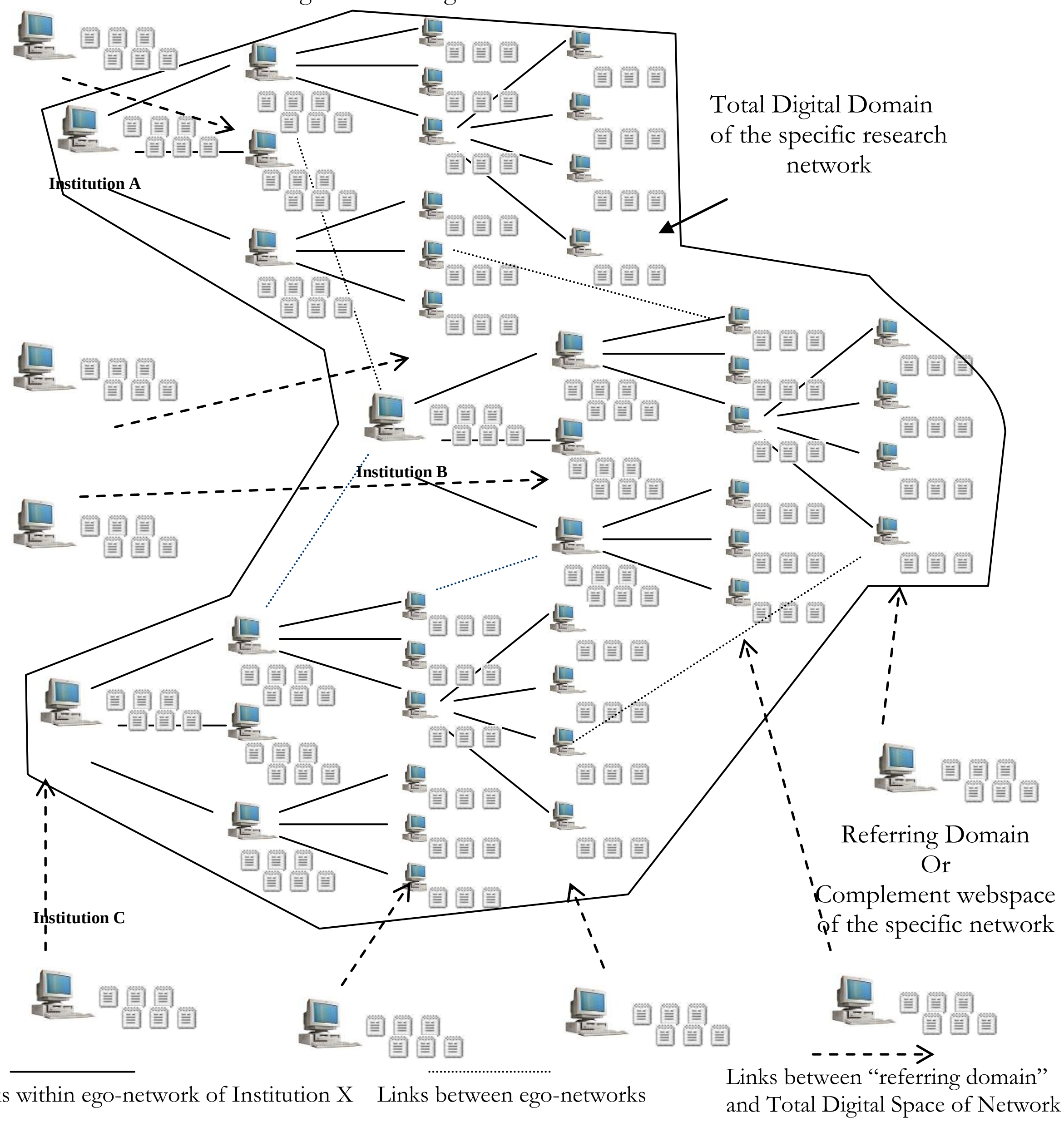
Two important characteristics of this methodology are worth commenting upon. First, the initial restriction to a set of collaborating institutions is crucial in terms of the precision of the analysis. By limiting the analysis to only research institutions that collaborate intensively, we can restrict the content nature of the results to knowledge resources relevant to the field of research under investigation. Secondly, the integration in the overall analysis of important "external" authoritative institutions/resources brings diversity to the initial network connections, as well as potentially allowing the extension of these digital knowledge bases to a broader spectrum not foreseen initally. This greater heterogeneity is likely to yield important results in terms of the quality of information that can be accessed from these digital knowledge-bases.

The conceptual model represented in Figure 5-3-4 allows the identification of the essential elements:

a) the electronic ego-networks for the best connected institutions (linkages to a depth *n* of connectivity);
b) the cross-linkages between the ego-networks of institutions at any "layer" level and electronic connectivity to external authoritative resources;
c) the inlinks coming from hubs, authorities and external internet resources to the best connected institutions;
d) the identification of network zones of intensive information resources - digital knowledge bases, resulting from the connectivity of the institutions.

An integrated representation of these concepts and processes is show in Figure 5-3-4.

Figure 5-3-4 - Identification of Digital Knowledge Bases



Notwithstanding the possible identification of these digital knowledge bases, there are two important limiting factors to be taken into account:

- There is no *refereeing* mechanism or scientific validation of the "knowledge" resources available electronically other than the self-organising process inherent in the institutional practice of disclosing information on electronic networks. This might represent a problem in terms of the quality of digital knowledge-base resources;
- This analysis is cross-sectional in nature and does not take account of the dynamic nature of these electronic systems. The structural representation of the electronic connectivity and the mapping of knowledge bases constitutes a "snapshot" at a certain period in time. But empirical evidence testifies to the changing nature of information published on the web, which might cause dramatic changes in these structural

representations. This problem can only be overcome by a longitudinal analysis of these networks.

5.3.8 Conclusions

Recent research has highlighted that the existence of a mechanism of "self-organisation" of large, distributed electronic networks, such as the Internet, allows the efficient identification of "web communities" in digital space. In this research we have focused our analysis on the *inner structure* of such "web communities", namely the connectivity patterns and centrality of certain institutions within these restricted networks of electronic interactivity. Moreover, as the analysis is particularly concerned with scientific communities and research networks, the sociological context embedding the collaboration structures is likely to have a fundamental weight in the overall system. The argument is that the electronic connectivity patterns "extend" existing, more or less well established collaboration patterns in the physical world.

We found that the pattern of electronic connectivity of institutions belonging to research networks does indeed show a remarkable similarity to physical, non-electronic collaboration structures. Moreover, the "best connected" group of institutions supports a kind of self-reinforcing mechanism in their inner connections, revealing "local proximity in cyberspace."

Secondly, we have detected heterogeneity in the *inner structure* of the electronic networks, even among the "best connected" entities. Therefore, there is a reduced subgroup of the "very best" connected institutions not only within the "internal" network under analysis, but also with important external connections to "authoritative" institutions and internet resources.

For institutions that have an open and public policy of information disclosure within these electronic networks (and this excludes most private companies), the extensive mapping of their digital ego-networks enables identification of important zones of information and knowledge resources - digital knowledge bases. These electronic resources (e.g. electronic publications, projects documentation, researcher information) are linked to the best connected institutions in the research network, as well as to important hubs and authorities in electronic webspace, as described above.

These results have important implications for science policy, particularly with regard to efforts to strengthen the networking of research resources and increase the availability of important knowledge resources in electronic networks, but also in terms of the organisation of research networks and their production and distribution of knowledge results.

A serious limitation to this investigation is the cross-sectional character of the investigation. A longitudinal analysis of the dynamics of these connectivity patterns, as well as of the corresponding dynamics of the digital knowledge bases, is an important line for further empirical investigation. A comparison of the same conceptual scheme in different research networks, for example in different fields of research, or in research networks resulting from different funding policies, would be another important area for further analysis. Discussion of webmetric indicators at the micro-level of single institutions, the meso-level of groups of institutions or the macro-level of whole networks is, likewise, a research topic in itself. Research directed to analysing the policy implications of the commercial practice of non-public disclosure of information in these electronic environments is a another substantial research topic as are the issues related to copyright on information being disclosed by university and governmental institutions.

The technological and sociological dimensions of the problem of public availability of knowledge resources in electronic networks will potentially advance side by side in order to facilitate the positive prospects of such communication systems.

6. Conclusions and Research Limitations

This chapter summarises the main theoretical and empirical findings of the investigation, then discusses some methodological and other limitations of the research and finally outlines some areas of investigation for further research. In the first section, we underline the main contribution of the thesis and synthesise the answers to our initial research question and theoretical hypotheses based upon the empirical evidence (section 6.1 - synthesis of research). Section 6.2 revises the proposed conceptual model for electronic research collaboration and section 6.3 summarises the main conclusions from the research. In discussing some of the limitations to the investigation, we analyse bibliometric and scientometric problems (section 6.4.1), the importance of expert knowledge in the assessment of collaboration networks (section 6.4.2), the replication of this study in the analysis of domains of science other than computational speech and language (section 6.4.3) and, finally, the need for long-term evidence on electronic patterns of communication (section 6.4.4). Based upon the results of the investigation and its limitations, in the final section of the chapter (section 6.5) we outline some complementary lines for research on the more general theme of e-science.

6.1 Synthesis of Research

This investigation has produced theoretical and empirical evidence on how electronic networks are being used as effective infrastructures supporting social networks of research communication and collaboration. This new insight into the electronic networking of science networks constitutes the main contribution to knowledge of this research. In order to gain a better understanding of *how* the electronic networking of research collaboration is occurring, this study has produced knowledge about how traditional (non-electronic) patterns of communication and collaboration are being reproduced in electronic environments. In the following paragraphs we discuss why the electronic networking of research is of considerable importance for science policy. We then summarise how the electronic networking of research could be operationalised as the focus of empirical and theoretical analysis. Finally, we revisit the initial research question in the light of our new empirical evidence. This synthesis opens the way to a reconceptualisation of the model for electronic research collaboration (discussed in the next section).

The electronic networking of science (e-science) and, particularly, of research communication and collaboration, is an important topic for science policy intervention. The potential of ICTs in general, and electronic networks (supported by Internet technologies) in particular, to influence the organisation of research is great. This impact is verifiable in terms of the composition of research communities, the extension of traditional forms of organisation and the emergence of new modes and patterns of organisation of research communities. These structural transformations are likely to have a significant impact on research productivity and on the advancement of knowledge. The realisation of this potential is strongly correlated with the intertwining of two complementary dimensions: the *technological infrastructure* and the whole set of capabilities in electronic networking; and the *socio-organisational infrastructure* - the characteristics of research communities, the particular features of processes of scientific communication and the nature of research collaboration.

Extensive research has concentrated on examining the characteristics of the technological infrastructure of science and also the socio-organisation of science *per se*, but much less effort has been focused on the socio-organisational nature of e-science systems. The theoretical and empirical evidence examined in this investigation contribute to filling this gap in the knowledge. Several important research issues have been tackled in previous research. For reasons of clarity and in order to establish a conceptual framework, we can classify these investigations as providing digital infrastructures for science in three complementary categories: facilitation of group-work in science; dissemination and inter-change of information and knowledge in science; and the electronic support of social networks in science. In the first category, the literature is extensive on the following kinds of analyses: the technological and physical networks supporting scientific work (e.g. the need for high-bandwidth and quality of service in scientific electronic network systems); the development of specialised middleware for scientific work (e.g. Grid architectures); and the development of special-purpose scientific applications (e.g. simulation tools, computer-aided software engineering). In the second category - information and knowledge dissemination - there is also a growing and extensive corpus of research on such issues as: remote access to scientific instrumentation (e.g. in remote equipment for physics and astronomy or remote stations in oceanography); remote access to databases (e.g. the international genome databases); and digital libraries research and electronic publishing in science (e.g. e-print archives versus traditional publishing, large biomedical digital libraries and, more recently, the Internet as a huge digital library system). In the final category - electronic science networks - several projects and investigations have also been carried out, particularly in terms of: the implementation of large-scale electronic research and education networks (e.g. Internet2 in the US, or the GigaEthernet project in Europe); and extensive research on Collaboratories, virtual laboratories and computer-supported cooperative work in general (e.g. collaboratories in physics and other fields, and computer-mediated communication investigations). In each of these three categories, much more emphasis has been given to the characteristics of the technological infrastructure than to the socio-organisation of these electronic science systems. This current investigation has examined empirical evidence on this latter dimension in order to reconceptualise our theoretical knowledge about electronic research collaboration.

The focus of the investigation has been directed to the analysis of changing patterns of communication and collaboration in these new e-science environments (particularly supported by Internet technologies). When one analyses to what extent these electronic networks have been used as effective systems of social interaction, research communication and collaboration, answers were sought to the following research questions.

What is the structure of electronic scientific communication? Is this structure reproducing traditional (non-electronic) patterns? If so, to what extent is this transformation occurring?

The empirical evidence (analysed extensively in chapters 4 and 5 for the computational speech and language research field) apparently confirms that traditional patterns and forms of communication and collaboration in science are indeed being reproduced in electronic infrastructures. As such, patterns of electronic collaboration in science are expected to have a similar structure to traditional forms and patterns of organisation - thus, the structural transformations in e-science are likely to be evolutionary rather than revolutionary.

Chapter 4 provided empirical evidence on particular features of communication and collaboration in the field of computational speech and language (based upon bibliometric and scientometric analyses - see section 4.1) and on patterns of organisation and networking (based upon extensive network analysis of research project co-operation and participation in institutionalised research networks - section 4.2). Moreover, the combination of bibliometric and network analyses allowed the delimitation of a boundary for a set of researchers in computational speech and language whose responses to the survey provided detailed information at the level of each individual researcher about the use of Internet technologies for research work, communication and collaboration.

Chapter 5 provided detailed empirical evidence about features and patterns of the use of electronic networks in supporting research communication and collaboration. Two complementary strategies were used for collecting information: on the one hand, individual researchers self-reporting their experience of using electronic networks for science collaboration; and, on the other hand, detailed observation of electronic collaborative systems particularly suited to interactive inter-personal collaboration (newsgroups) and large-scale dissemination of information (the analysis of Internet inter-linkages among the set of institutions whose researchers were involved in the survey).

Analysis of the empirical evidence provided new insights in three complementary dimensions: 1) how electronic networks effectively facilitate group work in science; 2) how electronic networks support research collaboration (social network interactions); and 3) how electronic networks support the dissemination and inter-change of information and knowledge.

With regard to the use of electronic network technologies for facilitation of groupwork, the empirical evidence confirms the intensive and frequency of use of technologies for personal use and interchange of information such as E-mail and web browsing, whether for local or remote communication (E-mail is rated as *Very Important* for Remote Communicaton, as opposed to Face-to-Face which is rated *Average*, in the survey). E-mail is still rated as *Important* for Local Communication being second after Face-to-Face in Local Communication (rated *Very Important*). The whole set of Internet technologies is supporting remote collaboration and hence geographically dispersed and distributed workgroups (whether for co-authoring publications, working in research projects or organising seminars and conferences – see empirical results in Table 5-1-II on page 178).

Section 5.2 has provided detailed quantitative and qualitative evidence of how Newsgroups (as a kind of interactive electronic environment) are extending "invisible colleges" into the digital infrastructure of science. Informal social networks and electronic links of information were also confirmed in the Newsgroup analysis as well as in the digital knowledge bases section (section 5.3). A reasonable proportion of researchers (49%, 21% and 51%, respectively) rates as Important the use of a variety of technologies (E-mail, Web and FTP) for exchanging research results with the scientific community. And these technologies are also rated as *Important* for exchanging results with close collaborators (81%, 24% and 32% of the respondents to the survey, respectively). The use of Remote Servers with specific resources for speech and language has also been confirmed by the empirical data At least 27% of the respondents regularly (weekly) use high-performance computing platforms for speech and language processing.

Finally, electronic networks are being widely used as repositories of information (whether in the form of personal and institutional homepages (about 75% of total population reported having these kinds of on-line content), whether for access to electronic publications in Print-Servers (53.8% of the researchers), or more generally to inter-link with other researchers and research groups (see the results of digital knowledge bases analysis, in section 5.3).

Table 6-1-I – Comparison of Traditional to Electronic forms of Communication and Collaboration – Final Synthesis

Traditional Research Communication and Collaboration	Electronic Research Communication and Collaboration	
Research Communication and Group-Work		DIGITAL KNOWLEDGE BASES (large scale and distribu-ted Knowledge structures in electronic networks)
- Highly localised and bounded workgroups	- Highly distributed but limited e-mail Bulletin Board's workgroups	
- Geographically localised workgroups	- Geographically remote and distributed workgroups	
- Specialised Workgroups	- Electronically specialised workgroups	
- Linear research process workgroup (from initial idea to project's results)	- "Complex systems" workgroup research GRID architectures, resources and tools	
Collaboration Structures and Social Networks		
- "Invisible Colleges" – close knit and tightly connected groups of peers	- Electronic Invisible Colleges – Newsgroup invisible colleges or E-Mail Peer's communities; Pre-Print Server's Communities	
- Research Groups and Department Groups	- Remote "extended research groups" and "mission-specific" electronic research groups	
- Informal Social Networks	- Newsgroups, electronic links among researchers and research groups, electronic "webs of assistance"	
- Institutionalised Research Networks	- Remote Servers with network specific resources; Collaboratories	
Information Dissemination and Distribution		
- Preliminary ideas and research proposals	- Personal e-mail documentation	
- Meeting reports and documentation	- Mailing lists and FTP or Remote Servers	
- Conference and Symposia reports, articles and proceedings	- Internet Conference tools and repositories of information, Digital Libraries	
- Research projects documentation	- Electronic research projects (with modularity*)	
- Personnel and institutional information	- Personal and Institutional web presence	
- Pre-print publications	- E-Print Archives (with modularity*)	
- Journal articles	- Electronic Journals and E-Publications (with modularity*)	
- Books	- Electronic Books (with modularity*)	

* Modularity allows the computation of parts and components of documents in various formats.

Some more specific and detailed empirical evidence is discussed in the following paragraphs.

From the analysis of patterns of individual researcher's use of Internet technologies for research work, communication and collaboration (and particularly based upon the results of the electronic survey in section 5.1), certain conclusions can be drawn. First, Internet technologies are being used to support "extended research groups" as they support remote collaboration as well as local work-group collaboration. Secondly, we found that Internet technologies are being used with different degrees of intensity and frequency (e.g. e-mail and web browsing are used frequently and very intensively, while e-print archives and remote servers for advanced processing are used less frequently and with less intensity). Thirdly, the empirical data corroborate the hypothesis that different technologies are being used for different aspects of research work, communication and collaboration. For example e-mail is used as an effective inter-personal collaboration tool, as well as a collaboration tool for more extended research group-work. E-mail is also likely to be used in informal as well as formal activities of communication. Newsgroups are essentially a collaborative tool, but the empirical research has shown the content being exchanged within these environments is of a more informal nature. Web browsing tools are mostly used for passive and non-interactive research activities and for the dissemination of information. Notwithstanding this multi-modality in patterns of use of different technologies for different activities in research communication and collaboration, in looking at the global use of the whole set of technologies a very wide spectrum of activities can be found (from the more informal to the extremely formal, and from the more inter-individual to highly collaborative research collaboration activities - see more about this in the following section). Fourthly, we found these electronic networks to be important infrastructures for the wide and public dissemination of information and knowledge resources (individual researchers' data, scientific publications, links to collaborators, etc.). This result is reinforced by the analysis of the role of electronic networks in supporting the emergence of digital knowledge bases (see the last paragraphs of this section). Finally, we found that specialised electronic network infrastructures for science - such as advanced networking and high-performance computing (for example, for speech and language processing, access to remote databases, more demanding network architectures for remote collaboration) are being used less intensively but nevertheless with some regularity. These results emphasise the need, in terms of science policy, for the strategic implementation of more advanced e-science infrastructures.

In focusing on empirical evidence for how effectively electronic networks reproduce traditional structures and patterns of scientific communication and collaboration (comparing patterns of research collaboration as revealed in chapter 4 with empirical evidence on electronic networking in chapter 5), we find confirmation along several dimensions: patterns of connectivity, specialisation, centrality and prestige, diversity of interaction and universality. First, it was found that the electronic networking of researchers and research groups reveals hierarchical structures similar to those in non-electronic forms of research organisation. Secondly, the long-term analysis of particularly interactive electronic collaboration environments - newsgroups - provided good evidence that the evolution of electronic collaboration systems follows a "specialisation" model, with an endogenous division of labour - also a typical feature of traditional scientific organisation and forms of research group collaboration. Thirdly, we found that highly skewed differences in the distribution of electronic prestige and in the centrality of certain researchers and reseach groups in electronic environments follow quite similar patterns to those in non-electronic science - e.g. Lotka's Law of scientific productivity, the Matthew effect and cumulative scientific recognition. Fourthly, the diversity of interactions found in scientific collaboration - and materialised in inter-sectoral and international networks of research communication and collaboration - is naturally reproduced in these electronic networks. Finally, the Mertonian norms of universality and communalism, expected to a certain degree in science, also find representation in these electronic networks, through the extension of remote interaction and the public and large-scale dissemination of information and knowledge.

Finally, from analysis of the extent to which these electronic networks are contributing to the large-scale dissemination of information and knowledge-sharing activities (see particularly section 5.3 of the empirical analysis), it can be seen that Internet technologies - and particularly the World Wide Web - are indeed contributing to the emergence of large-scale and distributed knowledge structures - digital knowledge bases. These digital knowledge bases are not merely repositories of scientific information resources – researchers' contacts, research project documentation, scientific databases, scientific tools, pre-prints, research group and institutional information, electronic publications and links to collaborating research groups and institutions. They also reproduce non-electronic patterns of connectivity, hierarchies in collaboration and differences in prestige and centrality, as well as extended links to authoritative research groups in the scientific community. For instance, detailed analysis of the topology of electronic networking - at the level of research groups - reveals a remarkable similarity with physical, non-electronic collaboration networks. In addition, the heterogeneity, hierarchical structure of scientific organisation and differences in centrality and prestige in electronic space correlate with non-electronic research characteristics. Also, the open and public dissemination of information in electronic networks is not the norm for all research groups and institutions. We found that this policy of open-sharing of information is apparently confined to public research institutions and particularly to universities. Private research laboratories do not usually contribute to public digital knowledge bases. Finally, electronic connectivity to important research agents (e.g. highly prestigious research centres in certain subspecialties) provides a long-range extension of the "digital knowledge base" to remote sources. The empirical confirmation of the existence of these "digital knowledge bases" leads to some science policy implications, some of which are briefly discussed in the conclusions (section 6.3).

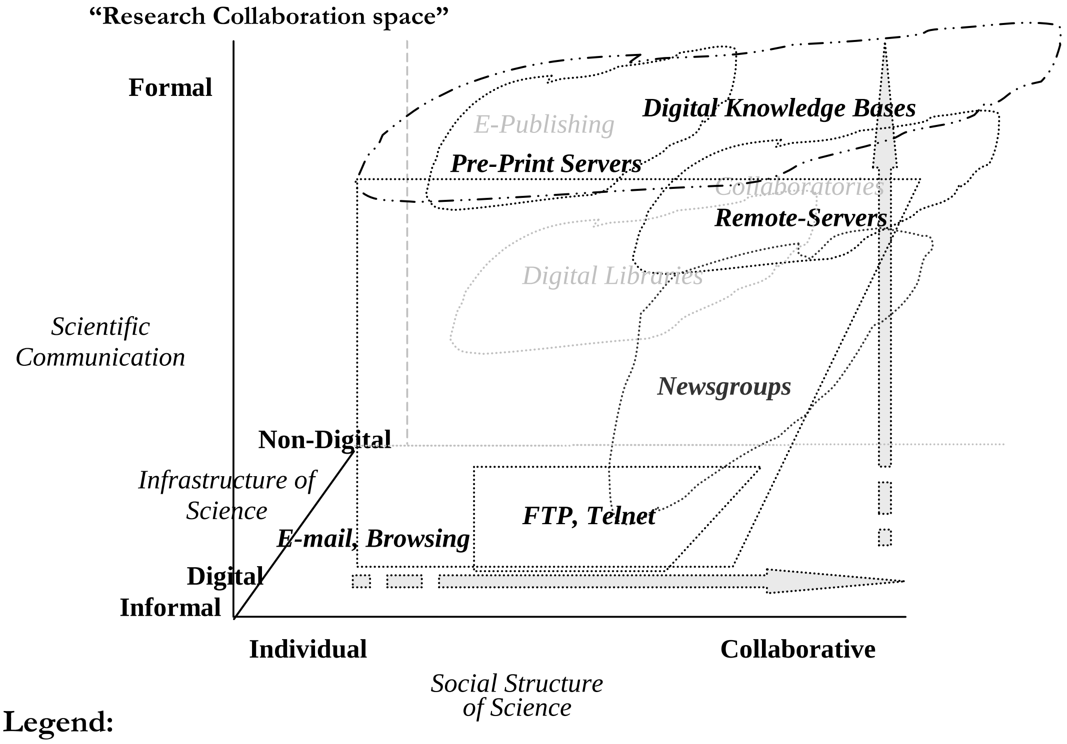
The integrated analysis of the above empirical evidence on how electronic research collaboration occurs provides a good basis for discussion of a revision to the initial conceptual model of electronic collaboration presented in this thesis.

6.2 An Extended Model of Electronic Scientific Communication. Digital Knowledge Bases

In the conceptual framework chapter (see section 2.2) a conceptual model was proposed, which is depicted in Figure 6-2-1. Based upon our empirical evidence on the use of electronic networks as systems for scientific communication and collaboration, we are now able to refine the initially proposed model to provide a better representation of "electronic research collaboration". This refinement is based upon empirical evidence on the area of computational speech and language analysed in terms of: a) the use of Internet technologies for research group-work, communication and collaboration - patterns of ICT use; b) the extent of reproduction of traditional forms of communication and collaboration - in other words, how electronic networks constitute effective electronic collaboration networks; and c) the extent of dissemination and interchange of information in these electronic networks, and the existence and extent of large-scale and distributed digital knowledge bases. Figure 6-2-1 provides a representation of this revised model for electronic research collaboration.

The investigation has provided good evidence for the analysis of electronic networks, and particularly digital infrastructures supported by Internet technologies, as electronic systems for social interaction, research communication and collaboration. Moreover, the empirical analysis apparently confirms that traditional patterns of communication and collaboration in science are being reproduced in these electronic networks. In this sense, the conceptualisation of electronic research collaboration as involving a spectrum of activities in three complementary dimensions (the formality of the communication, the collaborative nature of organisation of research and the nature of the technological infrastructure) seems to be strenghtened by this empirical analysis.

Figure 6-2-1 - "Electronic Research Collaboration" model



Legend:

E-mail, ... **Text in bold** - Technologies subject to empirical analysis in Speech and Language (E-mail, Browsing, FTP, Telnet, Newsgroups, Remote-Servers, Pre-Print Servers and Digital Knowledge Bases)

Digital Lib. **Text in Grey** - Technologies not subject to empirical analysis in Speech and Language (Digital Libraries, Collaboratories, E-Publishing)

Volumetric forms - Indicative of extension of applicability of each technology

Trend Arrows – Indicative of potential evolution in the use of technologies

The model serves two main purposes. First, it allows the graphical representation of the applicability of each Internet technology, based upon the empirical evidence analysed in the field of computational speech and language. Second, it serves as a more generic model to be applied in other fields, and also covering a more extensive set of Internet Technologies.

Based upon the analysis of the empirical evidence in speech and language, we can now elaborate and refine our initial conceptual maps and provide some more insigth to the more general applicability of these technologies.

First, we can now provide a more rigorous positioning of the applicability (actual use) of each technology in the spectrum of collaborative activities. We now have a better understanding of where the "border line" of applicability of Internet technologies lies. This advance is represented in the model by delimiting in a more rigorous way the "limits" of technology use. In this sense, the model conceptualises more thoroughly the extent of multi-modality of each technology - i.e. the same technology supporting a wider or narrower spectrum of activities; for example, e-mail and web browsing are now known to support a wide range of research communication and collaboration activities (from the informal to highly formalised forms of communication, and from the more individual to the highly collaborative interactions). On the other hand, FTP and Telnet technologies are now known to have a more restricted applicability. Nevertheless, the volumetric forms represented graphically in the model should be taken primarily as indicative, rather than as precise and rigorous delimitations of technology's applicability.

Secondly, given the empirical evidence on the frequency and intensity of use of different technologies for different stages and processes of research work, communication and collaboration, we can now conclude that at this point, electronic research collaboration is occurring more intensively in supporting more informal forms of communication and interaction, rather than supporting more formalised processes of communication in science. Likewise, electronic research collaboration is more extensive in inter-individual and inter-group activities, than in institutionalised collaborative work. In this sense, we could hypothesise that the evolution in electronic research collaboration is likely to be in the direction of more formalised forms of collaboration and more collaborative forms of electronic organisation. Both premises are depicted graphically in the model by two arrows revealing the trend of electronic networking support from individual to highly collaborative activities and from informal to more formalised processes of communication and collaboration.

Thirdly, given the empirical results, we are likely to define more rigorously the applicability of the technologies that facilitate group-work rather than technologies providing large-scale information dissemination and knowledge inter-change. In the same vein, this latter category of technologies is better identified than the technologies supporting electronic social networking. Again, this suggests evolution from the use of electronic networks as technological infrastructures to a wider use of these systems as socio-organisational electronic systems.

Fourthly, as a more generic model to be applied in other fields of research and in analyses of other kinds of technologies, we can use the same model to localise the use and extension of applicability of technologies such as E-Publishing, Digital Libraries or Collaboratories. Based upon previous evidence collected in the literature review (Crawford *et al.*, 1996; Walsh and Bayma, 1996) in other fields of research (*e.g.* Astronomy, High-Energy Physics, Molecular Biology) a preliminary potential and primarily indicative graphical representation of these technology's applicability is also provided in Figure 6-2-1.

Finally, and again as a more generic model to be applied in other fields of research, the model accounts for the dynamics in electronic research collaboration, both in terms of time-series analyses and in terms of cross-sectional analyses. With regard to structural evolution over time - that is, the evolution in the use of these technologies for research communication and collaboration by the same community of researchers or research groups - we can depict in the same model different patterns of use of each technology at different points in time, or alternatively, we can formulate two or more models, each representing the same community at different points in time. In comparing use of electronic networks for research collaboration in different communities, we can depict two or more models, each representing a different community.

The conceptual model for electronic research collaboration offers a very robust mechanism for identifying the extent to which electronic networks (and digital infrastructures in general) are supporting research communication and collaboration. The empirical analysis has not only strengthened its applicability, but also has provided a more rigourous delimitation of the combination of the whole set of technologies. The following section summarises the conclusions by revisiting the initial theoretical hypotheses based upon our empirical analyses.

6.3 Conclusions of the Research

The realisation of the potential of ICTs in general, and Internet technologies and electronic networks in particular, in terms of their positive influence upon science, is strongly dependent on the extent to which the new technological infrastructure effectively supports the socio-organisational practices of scientific communities. This investigation has produced new theoretical and empirical evidence on how these networks are being used as effective electronic systems of social interaction. In particular, we have focused our analysis on how changing patterns of communication and collaboration in science are being reproduced in electronic environments.

Adopting this structural perspective on the analysis of electronic networks, means they become more like social networks of interaction, communication and collaboration than technological infrastructures. From this perspective, it is important to analyse the adoption of electronic networks as systems for interaction among researchers, research groups and research institutions, for a wide spectrum of collaborative activities. This study has extended traditional analyses of technological infrastructures facilitating group work. Also, it is important to analyse electronic patterns of connectivity and to assess whether traditional forms of communication and collaboration are being extended into electronic environments. The study has extended previous work on collaborative tools for scientific work and contributes to our knowledge in the new area of examining electronic networks as electronic social networks. Finally, it is important to analyse the dissemination of information and the interchange of knowledge resources in the context of social networks of interaction. This study has extended our knowledge of how patterns of interaction, communication and collaboration, when complemented by open and public dissemination of scientific information, provide a basis for large-scale and distributed digital knowledge bases.

The empirical evidence analysed in this investigation has brought a new understanding of the extent to which traditional forms of organisation, communication and collaboration in science are being reproduced in electronic networks. Moreover, by focusing on patterns of electronic interaction, it has generated knowledge on patterns of use of electronic networks for research interaction (communication and collaboration), the detailed analysis of patterns of electronic connectivity (e.g. hierarchical structures, specialisation, organisation of electronic networks, diversity of forms of collaboration, evolution of electronic systems) and the extent to which they extend traditional (non-electronic) forms of organisation and, eventually, the emergence of new forms of large-scale dissemination of information and knowledge.

The empirical analysis demonstrated that, while there is some variability among individual researchers, there are regularities in the usage patterns of different technologies for different aspects of research work, communication and collaboration. Also, we found that different technologies are being used with different degrees of intensity and frequency for different aspects of the wide spectrum of communication and collaborative activity - we termed this multi-modality. In addition, we found that the extension of multi-modality varies significantly across technologies. The combination of this empirical evidence provides confirmation for our first theoretical hypothesis.
There is multi-modality in the use of Information and Communication Technologies for different stages of research work, communication and collaboration.

The empirical analysis also provided confirmation of our second theoretical hypothesis, concerning the extent to which electronic networks support traditional social networks of communication and collaboration.
Electronic scientific communication reproduces traditional structures of scientific communication and research collaboration.

We found that traditional features of scientific communication and the organisation of collaboration in science (such as the "Matthew effect" or "Lotka's Law", and the hierarchical structure of organisation and specialisation), typical of non-electronic scientific communication and collaboration, do indeed persist in electronic environments. Secondly, we found that patterns of connectivity in research collaboration, such as the significant variation in the centrality and prestige of key researchers and research groups, and the significant clustering of research communities (as identified in networks of citation, co-authorship networks, or networks of collaborative projects) also have an electronic equivalent. Thirdly, we found that the structural evolution of electronic networking replicates traditional processes of "division of labour" and intersectoral and international collaboration, known to hold in traditional scientific communities and to be typical of the development of fields of research. Finally, we found that the open and public exchange of information and knowledge, expected in public science more so than in other social communities, is also replicated in electronic environments.

The empirical analysis also corroborates, although in more exploratory terms, our third theoretical hypothesis concerning the emergence of large-scale and distributed electronic systems of knowledge dissemination.

Non-electronic structures of research collaboration are reproduced in electronic environments and we are able to discover and identify "Digital Knowledge Bases" - large-scale and distributed electronic knowledge structures.

We found that tightly bounded and closely collaborating research groups reproduce their connectivity interactions in electronic networks (see previous comments). This condition confirmed that "research communities" are also organised in "electronic communities", and it is likely that they have, at least at this point, identical forms of organisation. Secondly, we found that information and knowledge resources (e.g. scientific publications, research project documentation, human resources information, links to collaborating groups) are being openly and publicly disseminated in these electronic networks. This observation is valid for public science, but not for private research groups and institutions. Thirdly, we found that science policy inititatives directed to strenghtening the public availability of information and sharing of knowledge contribute to the open and public availability of information on these electronic networks.

This leads to some science policy orientations regarding the public availability of research information on electronic networks (more particularly), and with the organisation of public funding programmes that in a more or less orchestrated way might lead to increasing volumes of information resources distributed in electronic networks.

Some other science policy implications resulting from the empirical results on digital knowledge bases are worth commenting upon. First, the use of cybermetric methods to identify the electronic connectivity among research groups – and hence digital knowledge bases - might be more efficient than other traditional and more labour intensive methods – such as bibliometrics or scientometrics – which might be of great value for research evaluation purposes. Secondly, the distinctive features of a digital knowledge base as "organic and large-scale structures of electronic information" provides an efficient mechanism for assessing the dynamics of research communities, or at least of more or less self-organising research networks within research communities. This migh well be used as an efficient method for studying the organisation of research fields and the formation and evolution of research communities. Finally, locating digital knowledge bases in the widely dis-organised and dynamic web space, is probably a very efficient methid for "business intelligence" and information research activities.

To sum up, the investigation has generated theoretical and empirical knowledge on how electronic networks are being used as infrastructures to support research communication and collaboration. In particular, it has provided a detailed analysis of the nature of socio-organisational e-science systems, patterns of ICT use and group-work, the replication of traditional forms of science organisation in these electronic social networks, and the public availability of large-scale and distributed digital knowledge bases. This contribution to knowledge is naturally subject to several research limitations and there are a number of other research issues that remain to be investigated. These are discussed in the following sections.

6.4 Limitations of the Research

Any theoretical and empirical investigation is always subject to certain limitations (methodological and other), as well as being unable to cover certain complementary research issues that for various reasons cannot be fully analysed. There are several factors that might contribute to this, such as the time-frame of the investigation, the lack of availability of empirical data, the inaccessibility of key research sources, limits in terms of the financial and logistical resources available for the research. Other limitations follow from the methodological decisions and the general state-of-the-art in relation to the methods, techniques and instruments employed in the investigation. In our current investigation, there are four complementary areas that imposed limits and constrained the extensiveness of the investigation: limits deriving from bibliometric and scientometric analyses (discussed in section 6.4.1); restrictions on a more extensive assessment by experts of collaboration patterns in computational speech and language (section 6.4.2); the focus of analysis on the specific community of computational speech and language (section 6.4.3); and, finally the limited availability of empirical data on electronic scientific communication (discussed in section 6.4.4).

In this section we discuss these four types of limitation while in the next section we briefly summarise the main lines of research to be covered in subsequent research.

6.4.1 Overcoming the Scientometric Problems

Bibliometric and scientometric analyses were used in the investigation mainly for two complementary purposes: to identify researchers and research groups whose collaboration patterns could provide information about electronic scientific communication (to be collected by other means, such as the survey); and to identify particular features of scientific communication and to characterise patterns of research collaboration in the field of computational speech and language (such as the evolution of the field of research, division of labour and differences in research organisation of subcommunities).

Apart from the more general criticisms of bibliometric and scientometric analyses (which have been extensively discussed in the conceptual framework in chapter 2 and in the empirical discussion in section 4.1), some specific limitations of bibliometric work in inter-disciplinary and application-oriented fields of research like computational speech and language should be noted.

First, there are classification problems in the bibliometric processing of scientific publications from inter-disciplinary areas, particularly when it comes to subfield and subtopic classification. Secondly, in highly dynamic fields of research that have yet to reach maturity (and which do not yet have an established set of core journals) it is difficult to determine the "boundaries" for the bibliometric sources. This was apparent in speech and even more so in language research. Thirdly, technology-oriented fields may be subject to policies of secrecy or to delays in publication due to the reward mechanisms involved in private research activity. Fourthly, fields of research that are growing might be subject to much instability in terms of research orientation and, therefore, bibliometric methods alone may not be sufficient to identify all the key researchers and research groups or all the key topics for research.

The above constitute some of the main limitations encountered in the bibliometric analysis. The investigation attempted to combine a variety of techniques - in an extended broad search method - using core journal search, keyword search and citation of key-researchers in order to gain a clearer map of collaborative activity. In order to complement the bibliometric analysis, a network analysis of research activity was also carried out, along with a limited consultation with experts in the field.

6.4.2 The Need for "Expert" Validation of Collaboration Networks

In the course of the investigation, two experts in the field of computational speech and language provided information about the nature of the field of research, the evolution of speech and language, the most important journals and conferences in this field and important networking efforts in relation to collaboration in speech and language. Of particular value was the information about bibliometric references and about more or less institutionalised research networks. This provided a link to the more extensve analysis of research collaboration networks in European speech and language, which was extremely useful in conducting the electronic survey and the analysis of European funding of cooperative research projects. For reasons of time, as well as limits to accessing a wider network of "experts", the investigation could not include the use of experts to validate the results.

A more thorough expert validation of the collaboration networks identified by bibliometric and network analyses could have provided a more rigourous delimitation of the research community of speech and language. Since our principal focus was on determining the nature of electronic collaboration - and not on research evaluation - we adopted a more practical approach to delimiting the boundaries of the community in combining the results of the bibliometric and network analyses with information about institutionalised networks of research in European speech and language. However, the lack of an *interim* examination of collaboration networks, as well as the *ex-post* analysis of these collaboration networks, imposed some limits on the investigation in the initial stages of delimiting the boundaries of the community to be examined.

6.4.3 Replication of Research Strategy in other Fields of Research

The empirical evidence in this investigation comes from the analysis of patterns of communication and collaboration in the field of computational speech and language and is particularly focused on European researchers and research groups. These patterns of collaboration were assessed in traditional non-electronic as well as electronic environments. The choice of this field for empirical and theoretical investigation was explained in the methodology chapter (chapter 3) and is related to the availability of information about electronic and non-electronic collaboration activities, to the fact that this field is likely to have a considerable social impact in terms of new forms of communication and collaboration and to the networking efforts in this field of research related to various science policy initiatives. The following methodological reasons reinforced this choice of the speech and language community for gaining empirical knowledge about electronic networking of research collaboration. First, it was judged that this community was likely to make extensive use of ICTs in their research work, communication and collaboration and, hence, we should be able to identify patterns of ICT use in the community. Secondly, this community is engaged in an highly dynamic research field, working at the knowledge frontier, with a very intensive pattern of scientific communication and experiencing frequent restructuring of research groups and research institutions. Thirdly, in this community there is considerable "networking" of research actors - researchers, research groups and research institutions - and of research resources - information, specific speech and language computational tools, remote databases and so on. The combination of all the above factors provided good support for the choice of this community for empirical analysis.

There is no reason why the results of this investigation could not be generalised to other research fields, assuming the availability of appropriate empirical data. However, previous research discussed in the literature review section (section 2.1.6) has underlined the differences across fields of research in the adoption of ICTs for research communication and collaboration. Those earlier studies show that more geographically dispersed and interdependent communities and also research relying heavily on expensive and large experimental research settings (such as experimental physics), tend to benefit more from electronic communication. It has also been found that more application-oriented fields (such as chemistry and experimental biology), with more obvious commercial interests, are less likely to make the use of ICTs for informal scientific communication, depending much more on formal communication and patenting activities. In the light of these results, we cannot generalise about the patterns of electronic research communication across the whole edifice of science.

What is needed is an extension of the methods and research design of the current investigation to other research areas. The resulting cross-field comparisons would then help to validate the current results. While this is well beyond the scope of the current project (for reasons of resources and time) it would certainly constitute an interesting line for future research.

6.4.4 The Need for Longer-term Empirical Evidence on Electronic Scientific Communication

While electronic networking technology has been available for some time, Internet technologies, such as the World Wide Web or collaboration Internet tools, are much more recent (less than a decade or so). The same is true for technologies supporting distributed computing for the amazing improvement in computational power and storage capabilities, for the telecommunication support of collaborative work and for suitable graphical user interfaces. Similarly, even though the use of Internet technlogies for communication and collaboration within research communities has a longer tradition, it is only much more recently that these scientific communication patterns have begun to be confronted with the extension to an enlarged communication spectrum, involving society in general and businesses in particular. The novel capabilities of the technological infrastructure, as well as new forms of socio-organisation of e-science systems, represent an opportunity for new lines of research, but also a limitation in terms of the availability of empirical data for detailed long-term analysis.

The structural analyses carried out in the course of this investigation - to assess patterns of ICT adoption for research work and communication, to map patterns of electronic connectivity, and to map patterns in the electronic dissemination of information - would clearly benefit from an extended and longer-term analysis of the use of Internet technologies by the research community. This would require more resources and more empirical data to be available than was the case for this investigation, but would provide a robust comparison of the dynamics and evolution of these electronic systems of collaboration over time. We saw what would be the benefits of this in the analysis of the evolution of newgroups, where the availability of time-series data for about a decade of electronic interaction made it possible to carry out a detailed examination of the structural evolution of specialisation and of the division of labour and the dynamics of "electronic invisible colleges". Similar time-series and comparative analyses could be extremely rewarding in terms of surveying the same community of researchers at different points in time, as well as for assessing the variation and evolution in patterns of electronic connectivity in digital knowledge bases.

This brings us finally to a synthesis of some of the most interesting lines for further investigation.

6.5 Lines for Further Investigation

Several research issues, complementary to the topic of electronic networking of research collaboration, are not covered in this investigation but constitute important lines for further research. Among the research issues of theoretical and empirical significance, are the following: intellectual property rights (IPR) and ownership of electronic information and electronic resources; the evolution of electronic research collaboration systems, including new methodologies, new indicators and new conceptual models; comparison of observational and qualitative data of ICT use with computer simulation methods; and, finally, the integration of electronic research communication systems with other extended environments for collaboration, including social communities such as education networks.

Despite their undoubted importance, we have not dealt here with issues relating to intellectual property rights for scientific information and for other resources such as computational tools being exchanged within electronic networks of scientific communication and collaboration. While the ownership and authorship rights of research information disseminated in open and public electronic networks should be protected from unauthorised appropriation in order to support the long-term incentive structures motivating their creation, excessive protection schemes may block the wider dissemination and large-scale availability of these information resources. The complexity of the research issues surrounding this topic is enormous, but these problems certainly constitute an important area for further research.

Another interesting topic for investigation concerns the analysis of the dynamics and evolution of electronic systems of social interaction. This area involves research on new methodological strategies and the creation of new indicators for assessing network dynamics, as well as new conceptual models of the dynamics of electronic networks. New methods need to be developed to provide a more robust assessment of electronic interactions over longer time periods and supporting comparative analyses of empirical data from different fields. New indicators might provide a better characterisation of network interactions and dynamics, whether at the micro-level of inter-individual communication, at the meso-level of the comparative analysis of groups of electronic agents, or at the macro-level of analysis of entire electronic systems. Finally, new conceptual models of the dynamics of electronic networks might provide a better understanding of how these electronic systems change over time in the context of the growth and evolution of complex systems.

A further line of investigation focuses on methodological improvements for assessing electronic systems of social interaction. Recent advances in computer simulation methods allow the implementation of interesting computer models able to simulate information dissemination and knowledge inter-change, in the more general context of diffusion processes and complex systems organisation. A potentially fruitful extension to the methods developed in the current investigation (see cybermetrics methods in section 3.5) could involve a comparison of the observed patterns of ICT use for research communication and collaboration and patterns of connectivity within electronic systems, with computer simulation models. Such a comparison would provide experimental data for assessing the significance of certain variables. Among these variables, the significance of the number and diversity of researchers and research groups could be tested and also the centrality of key actors within these networks and the structural impact of science policy initiatives.

A final line of research worth mentioning here involves the extension of the focus from the relatively closed system of research communities and scientific communication to a wider audience of participants (such as education communities and science policy entities). The research system is certainly not a closed and hermetically sealed system of social interaction. On the contrary, the systemic transactions with other communities of interest are well recognised in the governance of science. Consequently, a better understanding of how electronic systems of research communication and collaboration interact with electronic educational networks or the involvement of science policy actors constitutes an obvious line for further empirical analyses.

The above research issues illustrate the reasons for continuing work in order to advance our knowledge about e-Science and the socio-organisational nature of e-Science in pursuit of a better understanding of the structure of electronic scientific communication.

7. Bibliographic References

Afuah, A. and N. Bahram (1995). "The Hypercube of Innovation", *Research Policy*, 24, 51 - 76

Albert, R., Jeong, H. and A-L. Barabási (1999). "Diameter of the World Wide Web". *Nature*, Vol. 401, 130 - 131

Agogino, A. (1999). "Visions for a Digital Library for Science, Mathematics, Engineering and Technology Education (SMETE)", *Proceedings of the Fourth ACM conference on Digital Libraries*, August 11-14, Berkeley, CA, USA

Antonelli, C. (1992). *The Economics of Information Networks*, Amsterdam, North Holland

Antonelli, C. (1994), "Localized Knowledge, Percolation Processes and Information Network", *Working Paper*, Department of Economics, University of Torino

Antonelli, C., Geuna, A. and W. E. Steinmueller (2000). "Information and Communication Technologies and the Production, Distribution and Use of Knowledge", *International Journal of Technology Management,* Vol. 20, Nos. 1/2, 2000, pp. 72-94

Aquino, S. and T. Mitchell (2001). "Search Engines Ready to Learn". *Technology Review Online*, April

Arora, A. and A. Gambardella (1994). "The Changing Technology of Technological Change: General and Abstract Knowledge and the Division of Innovative Labour", *Research Policy*, 23, 523 - 532

Arora, A., A. Gambardella, and E. Rullani (1998), 'Division of Labour and the Locus of Inventive Activity', Journal of Management and Governance, 1: 123-140

Arora, A., David, P. and A. Gambardella (1998). "Reputation and Competence in Publicly Funded Science: Estimating the Effects on Research Group Productivity", *Les Annales d'Economie et de Statistiques*, No. 49/50, Special Issue

Atkinson, M. (2001). *The UK e-Science Programme.* CANARIE's 7th Annual Advanced Networks Workshop: Thursday, November 29, 2001

Aubert, J. E. and V. Bayar (1999). "Maximising the Benefits of Information Technology for Science: Overview and Major Issues", *Science, Technology and Industry Review* No. 24 ,7 - 26

Arms, W..Y. (2000). *Digital Libraries*. Massachusetts , MIT Press

Axmann, H-P and S. P. Axmann (1999). A Global Research Infrastructure for the Global Research Village: The European Perspective. *Science, Technology and Industry*, Review No. 24, 30-48.

Balconi, M., S. Breschi and Francesco Lissoni (2002). "Networks of inventors and the location of university research: an exploration of Italian data.", *DRUID PhD Winter 2002 Conference*, Aalborg, January 17-19, 2002

Banos-Ruiz, R., Moreno-Bailon, R., Conteras-Jimenes, E. and J. Courtial (1999). "Structure and Dynamics of Scientific Networks. Part I: Fundamentals of the Quantitative Model of Translation", *Scientometrics*, Vol. 44, No. 2, 217 - 234

Banos-Ruiz, R., Moreno-Bailon, R., Conteras-Jimenes, E. and J. Courtial (1999). "Structure and Dynamics of Scientific Networks. Part II: The New Zipff's Law, The Clusters of Co-Citations and The Model of The Descriptor Presence", *Scientometrics*, Vol. 44, No. 2, 235 - 265

Byers, S. (2001). *Opportunity for all in a world of change. The Governments White Paper on enterprise, skills and innovation*. Department of Trade and Industry (DTI), 13 th February, 2001

Barnes, B. and D. Edge (1982). *Science in Context - Readings in the Sociology of Science*, London, Open University Press

Bar-Ilan, J. (1997). "The 'Mad Cow Disease', Usenet Newsgroups and Bibliometric Laws", *Scientometrics*, Vol. 39, No 1, 29 - 55.

Becchetti, C. and L. P. Ricotti (1999). *Speech Recognition. Theory and C++ Implementation*, Chichester, John Wiley & Sons

Bonitz, M. (1989). *Fundamental principles in Scientific Communication.* ZIID, Dresden

Bordons, M., A. Zulueta, A. Cabrero and S. Barrigón (1995). "Identifying Research teams with bibliometric tools", in *Proceedings of the Fifth Biennial Conference of the International Society for Scientometrics and Informetrics* (Learned Information Medford), pages 83 - 92

Bordons, M. and A. Zulueta (1997). "Comparison of Research Team Activity in Two Biomedical Fields", *Scientometrics*, 40 (3), 423 - 436

Brin, S. and L. Page (1998). "The Anatomy of a Large-Scale Hypertextual Web Search Engine", in *Proceedings of the Seventh International World Wide Web Conference*, Amsterdam, Elsevier, pages 107 - 117

Brooks, H. (1994). "The Relationship between Science and Technology*", Research Policy*, 23, 5, 477 - 486

Callon, M. (1980). "Struggles and Negotiations to define what is Problematic and what is not: the Sociology of Translation', in The Social Process of Scientific Investigation: Sociology of the Sciences Yearbook, 4, edited by Karin D. Knorr, Roger Krohn and Richard D. Whitley, 197-219, Dordrecht and Boston, Mass.: Reidel

Callon, M. (1988) La science et ses réseaux. Genèse et circulation des faits scientifiques Paris, La Découverte, 1988.

Callon, M. (1995). "Four Models for The Dynamics of Science", in S. Jasanoff, G. Markle, J. Petersen and T. Pinch (1995), *Handbook of Science and Technology Studies*, London, Sage, pp. 29 - 63

Callon, M. (1994). "Is Science a Public Good?", *Science, Technology & Human Values*, Vol. 19, No. 4, 395 - 424

Callon, M. and B. Latour (1992). "Don't Throw the Baby Out with the Bath School. A Reply to Collins and Yearley", in A. Pickering (ed.), *Science as Practice and Culture*, Chicago, University of Chicago Press, pages 343 – 368

Callon, M. and B. Latour (1981). "Unscrewing the Big Leviathan: How do Actors Macrostructure Reality, in K. Knorr and initial??Cicourel (eds) name of book, place and publsiher, , pages 277 - 303

Canada Ministry of Science (1998). "The Networks of Centres of Excellence Selection Committee Report", September, available electronically from the following URL: http://www.nce.gc.ca/pubs/reports/98/selection98_e.htm, on the 1st of July 2002

Carley, K. M. (1996). "Communicating New Ideas: Potential Impact of Information and Telecommunication Technology", *Technology in Society*, Vol. 18, No 2, 219 - 230.

Catlett, C. (2001). *Global Grid Forum Documents and Recommendations: Process and Requirements.* Technical Report. Argonne National Laboratory, Global Grid Forum , June 2001

Chaomei, C. and R. J. Paul (2001). "Visualising a Knowledge Domain's Intellectual Structure". *Computer*, IEEE, March, 65 - 71

Chollet, G., M. G. Di Benedetto, A. Esposito, M. Marinaro, 1998. Speech Processing, Recognition, and Artificial Neural Networks, Springer-Verlag Publisher.

Cohen, W.M., R. Nelson, J.P.Walsh (2002). "Links and Impacts: the Influence of Public Research on Industrial R&D", *Management Science*, 48, n.º 1, 2002, 1- 23

Cole, S and J. Cole (1967), "Scientific Output and Recognition", *American Sociological Review*, 32, 377 - 390

Cole, S and J. Cole (1973), *Social Stratification in Science*, Chicago, Chicago University Press

Cole, S. (1978) "Scientific Reward Systems: A comparative analysis", in R. A. Jones (ed.), *Research in Sociology of Knowledge, Sciences and Art*, Greenwich, Conn., JAI Press, pages 167 - 190

Cole, J. R. and H. Zuckerman (1975). "The Emergence of a Scientific Specialty: The Self-Exemplifying Case of the Sociology of Science". In TheldeaoJSocia/Srrucrure.L. A. Coser, Ed.: 139-174

Cole, R. A., Mariani, J., Uszkoreit, H., Zaenen, A. and V. Zue (1997). *Survey of the State of the Art in Human language Technology*. Cambridge, Cambridge University Press, HTML version available at http://cslu.cse.ogi.edu/HLTsurvey/, accessed 27 June 2002.

Collins, H. (1982). "Tacit Knowledge and Scientific Networks", in B. Barnes and D. Edge (eds) *Science in Context - Readings in the Sociology of Science*, London, Open University Press, pages 44 - 64

Collins, H and S. Yearley (1992). "Epistemological Chicken", in A. Pickering (ed.), *Science as Practice and Culture*, Chicago, University of Chicago Press, pages 301 - 326

Collins, H and S. Yearley (1992). "Journey into Space", in A. Pickering (ed.), *Science as Practice and Culture*, Chicago, University of Chicago Press, pages 369 - 389

Cooley, R., Mobasher, B., and J. Srivastava (1997). "Web Mining Information and Pattern Discovery on the World Wide Web"., Proceedings of the 9th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'97), November

Cowan, R. and D. Foray (1997). "The Economics of Codification and the Diffusion of Knowledge", *Industrial and Corporate Change*, Vol. 6, 3

Cowan R., Jonard N. (1999): "Network Structure and the Diffusion of Knowledge", Research Memorandum 99028, MERIT, Maastricht University

Cowan, R. and N. Jonard (2000). "The Dynamics of Collective Invention", Resarch Memorandum, 00-018, MERIT, Maastricht University

Crane, D. (1972). *Invisible Colleges*, Chicago, University of Chicago Press

Crawford, S. Y., Hurd, J. M. and A. C. Weller (1996). *From Print to Electronic. The Transformation of Scientific Communication*, Information Today, Inc, Medford, New Jersey, American Society for Information Science, ASIS Monograph Series.

Cunningham, S. and M. Dillon (1997). "Authorship Patterns in Information Systems"., *Scientometrics*, Vol. 39, No.º 1, 19 - 27

DANTE (2002). GÉANT - Gigabit European Academic Research Network. Available at http://www.dante.net/geant/about-geant.html, accessed 10th June 2002

Dasgupta, P. and P. A. David (1987). "Information Disclosure and the Economics of Science and Technology" in G. Feiwel (ed.), *Arrow and the Ascent of Modern Economic Theory*, New York, New York University Press, pages 519 - 542

Dasgupta, P. and P. A. David (1994). "Towards a New Economics of Science", *Research Policy*, 23, 487 - 521

David, P. (1993), "Knowledge, Property, and the System Dynamics of Technological Change", inSummers, L. H., Shah, S. (eds.), Proceedings of the World Bank Annual Conference on Development Economics 1992, Supplement to The World Bank Economic Review

David, P. (1995). "Stratification in Research Network Formation - a Sketch - Model of the World without the EU: Discussant's Comments", presented to the Conference on "New Research Findings: The Economics of Scientific and Technological Research in Europe", University of Urbino, February, 24-25

David, P. (1997). "Resource Allocation and the Global Science Infrastructure: International Cooperation and the Economics of Access to Large-Scale Research Facilities", paper presented to the "International Workshop on the Global Science System in Transition", IIASA 9, Luxenbourg, May 23-25

David, P. (1998). "Communication Norms and the Collective Performance of 'Invisible Colleges'", in Creation and Transfer of Knowledge: Institutions and Incentives, eds. G. Barba Navaretti, P. Dasgupta and K.G. Maler, Berlin, Springer Verlag (1998).

David, P. and Flemming, T. (1996), "Communications, Creativity and Research Network Dynamics: an Elementary Economic Model of Scientific Communities", paper presented to the International Conference on Creation and Transfer of Knowledge, (Castelgandolfo), September 1; (later revised for presentation to the All Souls College, Oxford, Seminar on The Economics of Information Diffusion May 1996)

David, P. A. and D. Foray (1994). "Percolation Structures, Markov Random Fields and the Economics of EDI Standards Diffusion", in G. Pogorel (ed.), *Global Telecommunication Strategies and Technological Changes*, Amsterdam, North-Holland

David, P. A. and D. Foray (1995). "Accessing and Expanding the Knowledge—Base in Science and Technology", *STI Review - Science, Technology and Industry*, no. 16: 13-68

David, P. A. and D. Foray (1996). "Information Distribution and the Growth of Economically valuable Knowledge: A Rationale for Technological Infra-structural Policies", in M. Teubal, D. Foray, M. Justman and E. Zuscovitch (Eds.), *Technological Infrastructure Policy: an International Perspective*, Dordrecht and London, Kluwer Academic

David, P., D. Foray and W. Steinmueller (1999) "The Research Network and the new economicsof science: from metaphors to organisational behaviours", in A. Gambardella and F.Malerba (eds.) The Organization of Economic Innovation in Europe, CambridgeAcademic Press, Cambridge

Diamond, A. (1996). "The Economics of Science". *Knowledge and Policy: The International Journal of Knowledge Transfer and Utilisation*, Vol.9, Nos. 2 & 3, 6 - 49

Donohue, T. (1996). "Impact of Information Systems on Engineering.", *Technology in Society*, Vol. 18, No.º 2, 127 - 135

DOE (2000). "The Facts of the Matter: Finding, Understanding and Using Information about our Physical World", in *Workshop Report on a Future Information Infrastructure for the Physical Sciences*. Hosted by the Department of Energy at the National Academy of Sciences, May 30-31

Eckman, R. and R. E. Quandt (1999). *Technology and Scholarly Communication.*, University of California Press, Berkeley and Los Angels

Edge, D. (1979). "Quantitative Measures of Communication in Science: A critical review". *History of Science*, XVII, 102 - 134

Galison, P. and D. J. Stump (eds), (1996). *The Disunity of Science. Boundaries, Contexts, and Power*, Stanford, California, Stanford University Press

Edge, D. (1995). "Reinventing the Wheel", in S. Jasanoff, G. Markle, J. Petersen and T. Pinch (1995), *Handbook of Science and Technology Studies*, London, Sage, pages 3 - 23

Edge, D. (1982). *Science in Context - Readings in the Sociology of Science*, London, Open University Press, pages 21 - 34

European Association of Aerospace Industries AECMA (2001). Aerospace within the European Research Area. Contributions to the Debate on the European Commission Initiative. Centres of Excellence, January

European Commission (2000). *Towards a European Research Area*. Communication from the Commission to the Council, the European Parliament, The Economic and Social Committee and The Committee of The Regions, Brussels, 18 January, COM (2000) 6

European Commission (2001). *How to Map Excellence in Research and Technological Development in Europe*. Commission Staff Working Paper, Brussels, 12 March, SEC(2001) 434

European Commission (2002). *Provision for Implementing Networks of Excellence*. Working document, Unit B.2, Research DG, European Commission, Second Edition: 7 May, available at the following URL: http://www.europa.eu.int/comm/research/fp6/networks-ip.html, on the 1st July 2002

European Science Foundation (2000). "European Science Foundation Response to the Communication of the Commission of the European Communities 'Towards a European Research Area'" - January

European Science Foundation (2000). "European Science Policy Briefing 7: The Need for High Bandwidth Computer-based Networking in Europe" (February)

Falkingham, L. and R. Reeves (1999). "Context Analysis - A Technique for Analysing Research in a Field, Applied to Literature on The Management of R&D at the Section Level", *Scientometrics*, Vol. 42, No. 2 (1999), 97 – 120

Faulkner, W. (1994). "Conceptualising Knowledge Used in Innovation: A Second Look at the Science-Technology Distinction and Industrial Innovation, Science, Technology & Human Values", Vol. 19, No. 4, 425 – 458

Faulkner, W. and J. Senker (1994). "Making Sense of Diversity: Public-Private Sector Research Linkages in Three Technologies", *Research Policy*, 23, 673 - 695

Faulkner, W. and J. Senker (1995). *Knowledge Frontiers*, Oxford, Oxford University Press

Finholt, T. A. and G. M. Olson (1997) "From Laboratories to Collaboratories: A new organizational form for scientific collaboration. *Psychological Science*, 9, 1, 28-36.

Flake, G., Lawrence, S. and G. Giles (2000). "Efficient Identification of Web Communities." *Proceedings of the Sixth International Conference on Knowledge Discovery and Data Mining (ACM SIGKDD-2000)* , Boston, MA, pages 150 - 160

Flake, G., Lawrence, S. Giles, C. and F. Coetzee (2002) "Self-Organisation and Identification of Web Communities.", *Computer Magazine*, March, 66 - 71

Fleck, J. (1994). "Learning by Trying: The Implementation of Configurational Technology", *Research Policy*, 23, 637 - 652

Foray, D. (1998). "Extending the Knowledge Stock. Proceedings of the CRIS 1998 Conference", March, Luxembourg

Freeman, C. (1991). "Networks of Innovators: A Synthesis of Research Issues", *Research Policy*, 20, 499 - 514

Fujimura, J. (1992). "Crafting Science: Standardised Packages, Boundary Objects, and 'Translation'", in A. Pickering (ed..), *Science as Culture and Practice*, Chicago, University of Chicago Press, pages 168 - 211

Galison, P. (1997). *Image & Logic. A Material Culture of Microphysics*, Chicago, University of Chicago Press

Galison, P. (1996). Introduction: The Context of Disunity", in P. Galison and D. J. Stump (ed.), , *The Disunity of Science. Boundaries, Contexts, and Power*, Stanford, California, Stanford University Press, pages 1 - 33

Galison, P. (1996). "Computer Simulations and The Trading Zone", in P. Galison, P. and D. J. Stump (eds), *The Disunity of Science. Boundaries, Contexts, and Power*, Stanford, California, Stanford University Press, pages 118 - 157

Galison, P. and D. J. Stump (ed.), (1996). *The Disunity of Science. Boundaries, Contexts, and Power*, Stanford, California, Stanford University Press

Garcia-Fontes, W. and A. Geuna (1999). 'The Dynamic of Research Networks in Europe', in A. Gambardella and F. Malerba (eds.), The Organisation of Economic Innovation in Europe, Cambridge: Cambridge University Press, pp. 343-366

Garfield, E. (1979). *Citation Indexing: Its theory and Application in Science*. New York, Wiley

Garfield, E., I.H. Sher and R.J. Torpie (1964). *The use of citation data for writing the history of science.* Philadelphia: Institute for Scientific Information, 1964

Garton, L., Haythornthwaite, C. and B. Wellman (1999). "*Studying On-line Social Networks. In Doing Internet Research",* S. Jones (ed.), Thousand Oaks, CA., Sage

Garvey, W.D. (1979). *Communication: the essence of science - Facilitating information exchange among librarians, scientists, engineers and students.*, Pergamon Press, Oxford, 1979

Garvey, W.D.and B.C. Griffith (1972). "Communication and information processing within scientific disciplines: empirical findings for psychology". *Information Storage and Retrieval*, Vol. 8:123-126, 1972.

Garvey, W. D., Lin, N. and C. E. Nelson (1970). "Communication in the Physical and the Social Sciences". *Science*, Vol. 170, December, 1166 - 1173

Garvey, W. D. and B. C. Griffith (1968) "Studies of Social Innovations in Scientific Communication in Psychology". *American Psychologist*, Vol. 21, 1019 - 1036

Geuna, A. (1997). *The Economics of Knowledge Production,* Cheltenham, Edward Elgar Publishing

Geuna, A. (1999). "Patterns of University Research in Europe", in A. Gambardella and F. Malerba (eds.), The Organisation of Economic Innovation in Europe, Cambridge: Cambridge University Press, pp. 367-389, 1999

Gibbons, M. , Limoges, C.,Nowotny, H.,Schwartzman, S., Scott P. and M. Trow (1994). *The New Production of Knowledge. The Dynamics of Science and Research in Contemporary Societies*, London, Sage

Gilbert, G. N. and S. Woolgar (1974). "The Quantitative Study of Science: An Examination of the Literature". *Science Studies*, 4, 279 - 294

Gilbert, L. E. (1998). "Disciplinary Breadth and Interdisciplinary Knowledge Production". *Knowledge, Technology and Policy*, Vol. 11, Nos 1 & 2, 4 - 15

Ginsparg, P. (1994) "First Steps Towards Electronic Research Communication", *Computers in Physics* Vol. 8, No. 4, 390

Ginsparg, P. (1997). "The Future of Research Communication" in Proceedings of the Workshop The Impact of Electronic Publishing on the Academic Community, April 17-20, , Portland Press, Stockholm, 1997

Gittelman, M. and B. Kogut (2003). "Does good science lead to valuable knowledge? Biotechnology firms and the evolutionary logic of citation patterns", *Management Science*, 49, n.º 4, April 2003, 366 - 382

Glanzel, W., Schubert, A. and H. Czerwon (1999). "A Bibliometric Analysis of International Scientific Cooperation of the European Union (1985 - 1995)", *Scientometrics*, Vol. 45, 185 - 202

Godin, B. (1996). "Research and the Practice of Publication in Industries.", *Research Policy*, Vol. 25,, 587 - 606

Gooding, D. (1992). "Putting Agency back into Experiment", in A. Pickering (ed.), *Science as Culture and Practice*, Chicago, University of Chicago Press, pp. 65 - 112

Gould, S. J. (1985). *Ontogeny and Philogeny*. Harvard, Harvard University Press

Gould, S. J. (2001). *The Lying Stones of Marrakech. Penultimate Reflections in Natural History*, London, Vintage

Gould, S. J. (2002). *The Structure of Evolutionary Theory*. Cambridge MA, Harvard University Press

Granovetter, M. (1973). "The Strenght of Weak Ties.", in *American Journal of Sociology*, Vol. 78, pages 1360 – 1380

Griffith, B. C. (1989). "Understanding Science. Studies of Communication and Information". *Communication Research*, Vol. 16, No. 5, 600 - 614

Griffith, B. C. and N.C. Mullins (1972). "Coherent Social Groups in Scientific Change. "Invisible colleges" may be consistent throughout science.", *Science,* Vol. 177, 959 - 964

Gupta, B., Kumar. S and C. Karisiddappa (1997). "Collaboration Profile of Theoretical Population Genetics Specialty", *Scientometrics*, Vol. 39, No. 3 (1997), 293 - 314

Katz, S. and B. Martin (1997). "What is Research Collaboration?", *Research Policy*, 26, 1 - 18

Hacking, I. (1992). "The Self-Vindication of the Laboratory Sciences", in A. Pickering (Ed..), *Science as Culture and Practice*, Chicago, University of Chicago Press, pages 29 - 64

Hacking, I. (1996). "The Disunities of The Sciences", ", in P. Galison and D. J. Stump (ed.), (1996). *The Disunity of Science. Boundaries, Contexts, and Power*, Stanford, California, Stanford University Press, pages 37 - 74

Hagstrom, W. O. (1965). *The Scientific Community*, New York, Basic Books

Hagstrom, W. O. (1982). "Gift Giving as an Organisational Principle in Science," In: Barry Barnes and David Edge, Science in Context: readings in the sociology of science. Milton Keynes: The Open University

Haan, J. De (1997) "Authorship patterns in Dutch Sociology". *Scientometrics*, Vol. 39, No. 2 (1997), 197 - 208

Heming, R. (1996). "Impact of Information Technology on the Earth Sciences.", *Technology in Society*, Vol. 18, No. 3, 297 - 319

Hicks, D. and S. Katz (1995). "A Profile of British Science: An Overview of Comprehensive, Time Series, Bibliometric Data", SPRU *Working Paper*, University of Sussex

Hicks, D. and S. Katz (1996). "Hospitals: the hidden research system." *Science and Public Policy*, Vol. 23, No. 5, October 1996, pp. 297-304

Hicks, D. and S. Katz (1997). "A National Research Network viewed from an Industrial Perspective", Revue d' Economie Industrielle, No. 79, 1st Trimestre 1997, pp. 129-142.

Howels, J. (1995). "Going Global: the use of ICT networks in Research and Development", *Research Policy*, 24, 169 - 184

Huang, L. (2000), *A Survey On Web Information Retrieval Technologies*. State University of New York at Stony Brook. Accessed via HTTP June 6, 2002 at http://citeseer.nj.nec.com/336617.html

Huberman, B., Pirolli, P. Pitkow, J. and R. Lukose (1998). "Strong Regularities in World Wide Web Surfing". *Science*, Vol. 280, 95 - 97

IEEE Computer Magazine (1996). *Special Issue on Digital Libraries*. Institute of Electric and Electronic Engineers (IEEE), Computer Magazine, May 1996

Interactions (1998). *Special Issue on Collaboratories*. Association for Computing Machinery (ACM), Interactions, vol. 5, no. 3 (May), 1998

Jaffe, A., M. Trajtenberg, R. Henderson (1993). Geographic location of knowledge spillovers as evidenced by patent citations. *Quarterly Journal of Economics*, 108 (3), 577-598

Kircz, J. G. (1998). « Modularity: The Next Form of Scientific Information Presentation*", Journal of Documenntation,* 54, 2, 210 - 235

Katz, S. and B. Martin (1995). "What is Research Collaboration?", *Research Policy*, 26, 1 - 18

Katz, S. and D. Hicks (1994). "The Classification of Inter-Disciplinary Journals: A New Approach", paper presented at Fifth International Conference on 'Scientometrics and Informatics'", River Forest, I11

Katz, S. and D. Hicks (1997). Bibliometric indicators for national systems of innovation. Prepared for IDEA Project funded by the EC TSER Programme. Brighton: SPRU, August 1997

Kleinberg, J. (1998). "Authoritative Sources in a Hyperlinked Environment". *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms*, Proc. 9th ACM-SIAM Symposium on Discrete Algorithms, 1998. Extended version in Journal of the ACM 46(1999)

Kleinberg, J. and S. Lawrence (2001). "The Structure of the Web", *Science*, Vol. 294, 1849 - 1850

Kline, S. J. and N. Rosenberg (1986). "An Overview of Innovation" in R. Landau and N. Rosenberg (eds), *The Positive Sum-Strategy: Harnessing Technology for Economic Growth,* Washington DC, National Academy Press, pages 275 - 306

Kling, R. and G. McKim (1997). "Electronic Media and Scientific Communication: Opportunities, Complexities and Limitations", consultancy paper for OECD, Indiana University School of Library and Information Science, September, mimeo.

Knorr-Cetina, K (1992). "The Couch, The Cathedral, and the Laboratory: On the Relationship between Experiment and Laboratory in Science", in A. Pickering (ed..), *Science as Culture and Practice*, Chicago, University of Chicago Press, pages 113 - 138

Knorr-Cetina, K (1995). "Laboratory Studies: The Cultural Approach to the Study of Science", in S. Jasanoff, G. Markle, J. Petersen and T. Pinch (1995), *Handbook of Science and Technology Studies*, London, Sage, pages 140 - 166

Knorr-Cetina, K (1996). "The Care of the Self and Blind Variation: The Disunity of Two Leading Sciences", in P. Galison and D. Stump, *The Disunity of Science. Boundaries, Contexts and Power*, Stanford, CA, Stanford University Press, pages 287 - 310

Koenraad, D. and M. A. Rappa (1995). "Scientists at Major and Minor Universities: Mobility along the Prestige Continuum", *Research Policy*, 24, 137 - 150

Koku, E., Nazer, N. and B. Wellman (2000). Netting Scholars: Online and Offline. *American Behavioral Scientist*, Volume 43, Special Issue: "Mapping Globalisation"

Kosala, R. and H. Blockeel *(2000). "Web Mining Research: A Survey". Proceeding of ACM - SIGKDD Explorations*, *July 2000*, Vol. 2 Issue 1, 1- 15

Kuhn, T. (1962). *The Structure of Scientific Revolutions*, Chicago, University of Chicago Press

Kumar, R., Raghavan, P., Rajagopalan, S. and D. Sivakumar (2000). "The Web as a Graph". In Proceedings of the Nineteenth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, May 15-17, 2000

Kumar, R., Raghavan, P., Rajagopalan, S. and A. Tomkins (1999). "Extracting Large-scale Knowledge Bases from the Web". *Proceedings of the 25th VLDB Conference*, Edinburgh

Laface, P. and R. De Mori (eds) (1992). "Speech Recognition and Understanding. Recent Advances, Trends and Applications*"*, in NATO ASI Series, *Series F: Computer and Systems Sciences*, Vol. 75, , Berlin, Springer-Verlag

Larson, R. (1996). "Bibliometrics of the World Wide Web: An exploratory analysis of the intellectual structure of Cyberspace". Presented at the Annual Meeting of the American Society for Information Sciences, Baltimore, April 25-26

Latour, B. (1987*). Science in Action: How to Follow Scientists and Engineers through Society*, Cambridge MA, Harvard University Press

Latour, B. (1988*). The Pasteurisation of France*, Cambridge MA, Harvard University Press

Latour, B. and S. Woolgar (1982). "The Cycle of Credibility", in B. Barnes and D. Edge *Science in Context - Readings in the Sociology of Science*, London, Open University Press, pages 35 - 43

Lawrence, S., Bollacker, K. and C. L. Giles (1999). "Digital Libraries and Autonomous Citation Indexing", *IEEE Computer*, Vol. 32, No. 6, 67-71

Lawrence, S. and C. Giles (1999). "Accessibility of Information on the Web". *Nature*, Vol. 400 No. 6740, 107 - 109

Leeuwen, T. van and R. Tijssen (2000). "Interdisciplinary Dynamics of Modern Science: Analysis of Cross-disciplinary Citation Flows". *Research Evaluation*, Vol. 9, No. 3, 183 - 187

Leiner, B. M., Cerf, V.G., Clark, D.D. Kahn, R.E.. Kleinrock, L., Lynch, D.C., Postel, J., Roberts, L.G. and S. Wolff (eds) (2002). *A Brief History of the Internet* Internet Society, available at http://www.isoc.org/internet/history/brief.shtml, last updated Monday, 04-Mar-2002 08:40:36 EST

Leydesdorff, L. (2001). *A Sociological Theory of Communication: The Self-Organisation of the Knowledge-Based Society*, Universal Publishers, USA

Lewenstein, B.V. (1995). "Do Public Electronic Bulletin Boards Help Create Scientific Knowledge? The Cold Fusion Case", *Science, Technology and Human Values*, Vol. 20 No, 123 - 149

Licklider, J. C. R. (1968). "A Crux in Scientific and Technical Communication", *American Psychologist* 21, 1044 - 1051

Lievrouw, L. A. (1998), "Four Programs of Research in Scientific Communication". *Knowledge in Society*, Vol 1, N.o 2, 6 - 22

Lomas, R. (2002). *The Invisible College. The Royal Society, Freemasonry and the Birth of Modern Science*, London, Headline Book Publishing

Lotka, A.J. (1926), "The frequency distribution of scientific productivity.", *Journal of the Washington Academy of Sciences*, Vol. 16, page 317

Luukkonen, T. and S. Hälikkä (2000). *Knowledge Creation and Knowledge Diffusion Networks. Impacts in Finland of the EU's Fourth Framework Programme for Research and Development. TEKES*, Publications of the Finnish Secretariat for EU R&D 1/2000, Helsinki, Finnish Secretariat

Machlup, F. (1980). *Knowledge and Knowledge Production, Volume* I in F. Machlup, Knowledge: Its Creation, Distribution and Economic Significance. Guildford, Surrey: Princeton University Press

Machlup, F. (1982). The Branches of Learning, Volume II in Fritz Machlup, Knowledge: It's creation, distribution and Economic Significance. Guildford, Surrey, Princeton University Press

Machlup, F. (1984). The Economics of Information and Human Capital, Volume III in Fritz Machlup, Knowledge: It's creation, distribution and Economic Significance. Guildford, Surrey, Princeton University Press

MacRoberts, M. H. and B. R. MacRoberts (1982). "A Re-evaluation of Lotka's Law of Scientific Productivity". *Social Studies of Science*, Vol. 12, 443 - 450

MacRoberts, M. H. and B. R. MacRoberts (1986). "Quantitative Measures of Communication in Science: A Study of the Formal Level". *Social Studies of Science,* Vol. 16, 151 - 172

Martin, B. and J. F. Skea (1992). "Academic Research Performance Indicators: An Assessment of The Possibilities". Report to The Advisory Board for the Research Councils and the Economic and Social Research Council, Brighton, SPRU, University of Sussex

Martin, B., Skea, J.F. and E. Ling (1992). "Performance Indicators for Academic Scientific Research" End-of-award Report Project no. A418254009 submitted to the Advisory Board for The Research Councils and the Economic and Social Research Council, Brighton,. SPRU, University of Sussex

Martin, B. and J. Irvine (1989). *Research Foresight: Priority Setting in Science*, London, Pinter

Martin, B., Irvine, J. Narin, J. Sterritt, C. and K. A. Stevens (1990). "Recent Trends in the Output and Impact of British Science", *Science and Public Policy*, Vol. 17, 14 - 26

Meadows, A. J. (1998). *Communicating Research*, San Diego: Academic Press, 1998

Melin, G and O. Persson (1996). "Studying Research Collaboration using Co-authorships". *Scientometrics*, Vol. 36, 363 - 377

Meyer, M. and O. Persson (1998). "Nanotechnology - Interdisciplinarity, Patterns of Collaboration and Differences in Application". *Scientometrics*, Vol. 42, No. 2, 195 - 205

Menéndez, L-S, Bordon, M, and M. A. Zulueta (2001). "Interdisciplinarity as a Multidimensional Concept: its Measure in Three Different Research Areas". *Research Evaluation*, Vol. 10, No. 1, 47 - 58

Menzel, H. (1968). "Scientific Communication: Five themes from Social Science Research". *American Psychologist*, Vol. 21, 999 - 1010

Merton, R. (1942). "The Normative Structure of Science", in N. W. Storer (ed.), *The Sociology of Science. Theoretical and Empirical Investigations*, Chicago, University of Chicago Press, pages 267 - 278

Merton, R. (1957). "Priorities in Scientific Discovery", in N. W. Storer (ed.), *The Sociology of Science. Theoretical and Empirical Investigations*, Chicago, University of Chicago Press, pages 286 - 324

Merton, R. (1961). "Singletons and Multiples in Science", in N. W. Storer (ed.), *The Sociology of Science. Theoretical and Empirical Investigations*, Chicago, University of Chicago Press, pages 343 - 370

Merton, R. (1963). "Multiple Discoveries as Strategic Research Site", in N. W. Storer (ed.), *The Sociology of Science. Theoretical and Empirical Investigations*, Chicago, University of Chicago Press, pages 371 - 382

Merton, R. (1968a). "Behavior Patterns of Scientists", in N. W. Storer (ed.), *The Sociology of Science. Theoretical and Empirical Investigations*, Chicago, University of Chicago Press, pages 325 - 342

Merton, R. (1968b), "The Matthew Effect in Science. The reward and communication systems of science are considered.", *Science*, Vol. 159, 56 - 63

Merton, R. (1970). "Social and Cultural Contexts of Science", in N. W. Storer (ed.), *The Sociology of Science. Theoretical and Empirical Investigations,* Chicago, University of Chicago Press, pages 173 - 190

Moreno, J. L. (1932) *Group Methods and Group Psychotherapy*, Sociometry Monograph No5, Beacon: Beacon House

Mullins, N. C. (1972). "The Development of a Scientific Specialty: The Phage group and the origins of molecular biology." *Minerva*, Vol X, 3 - 29

Mullins, N. C., Lowell L. H., Hecht, P. K. and E. L. Kick (1977). "The Group Structure of Cocitation Clusters: A comparative study", *American Sociological Review*, 1977, Vol. 42. August, 552 - 562

National Research Council (1994), *Realizing the Information Future*, Washington, DC, National Academy Press.

National Research Council (1997), *Bits of Power*, Washington, DC, National Academy Press.

Newman, M. E. J (2001). "Who is the Best Connected Scientist? A study of scientific co-authorship networks". *Physical Review Letter*, available at http://www.arxiv.org/PS_cache/cond-mat/pdf/0011/0011144.pdf, March 2003

Noyons, E. C. M., Moed, H. F. and A. F. J. Van Raan (1999). "Integrating Research Performance Analysis and Science Mapping". *Scientometrics*, Vol. 46, No. 3, 591 - 604

Norwegian Ministry of Education, Research and Church Affairs (2000). "Norwegian views on the Commission Communication 'Towards a European research area'", Technical Report, reference 00/8890

OECD (1999). *Communications Outlook*, Paris. OECD

OECD (1998). *"The Global Research Village"*, available electronically at the following URL: http://www.oecd.org/dsti/sti/s_t/scs/prod/online.htm, last accessed 1st July 2002

Olson, G.M., and Olson, J.S. (2001) Distance matters. *Human-Computer Interaction*, 15, 139-179.

Patel, P. and K. Pavitt (1994). "Patterns of Technological Activity: Their Measurement and Interpretation", in P. Stoneman (Ed*), Handbook of the Economics of Innovation and Technological Change*, 1994

Pavitt, K. (1984). "Sectoral Patterns of Technical Change: Towards a Taxonomy and a Theory", *Research Policy*, 13, 343 - 373

Pavitt, K. (1991). "What makes Basic Research Economically Useful?", *Research Policy*, 20, 109 - 119

Pavitt, K. (1988). "Uses and Abuses of Patent Statistics", in A. F. J. van Raan (Ed.), *Handbook of Quantitative Studies in Science and Technology*, Elsevier Publishers

Persson, O., Melin, G., Danell, R. and A. Kaloudis (1997). "Research Collaboration at Nordic Universities". *Scientometrics*, Vol. 39, No. 2, 209 - 223

Persson, O. and M. Beckmann (1995). "Locating the Network of Interacting Authors in Scientific Specialties". *Scientometrics*, Volume 33, No 3, 351 - 366

Pickering, A. (1992). "From Science as Knowledge to Science as Practice", in A. Pickering (Ed..), *Science as Culture and Practice*, Chicago, University of Chicago Press, pages 1 - 26

Pickering, A. and Adam Stephanides (1992). "Constructing Quartenions: On the Analysis of Conceptual Practice", in A. Pickering (Ed..), *Science as Culture and Practice*, Chicago, University of Chicago Press, pages 139 - 167

Pickering, A., (1992). *Science as Practice and Culture*, Chicago Ill., University of Chicago Press

PITAC (1998). President's Information Technology Advisory Committee Interim Report to the President, National Coordination Office for Information Technology Research and Development, USA: Arlington, August 1998

PITAC (1999). President's Information Technology Advisory Committee, "Information Technology Research: Investing in Our Future", National Coordination Office for Information Technology Research and Development, USA: Arlington, February 1999

PITAC (2000). President's Information Technology Advisory Committee, "Resolving the Digital Divide: Information, Access and Opportunity", National Coordination Office for Information Technology Research and Development, USA: Arlington, February 2000

PITAC (2001). President's Information Technology Advisory Committee, "Digital Libraries: Universal Access to Human Knowledge", National Coordination Office for Information Technology Research and Development, USA: Arlington, February 2001

Polanyi, M. (1963). *Personal Knowledge: Towards a Post-Critical Philosophy*, Chicago Ill, University of Chicago Press

Price, D. de Solla (1986). *Little Science, Big Science … and Beyond*, New York, Columbia University Press

Price, D. de Solla (1981). "The Development and Structure of the Biomedical Literature", in K. S. Warren (ed.) *Coping with the Biomedical Literature: A Primer for the Scientist and the Clinician*, , New York, Praeger, 3 - 16

Price, D. de Solla (1980). "On the Scientific Element in a Scientific Communication". *Inerciencia*, Vol. 5, No. 4, 220 - 222

Price, D. de Solla and Donald de B. Beaver (1968). "Collaboration in an Invisible College". *American Psychologist*, Vol. 21, 1011 - 1018

Price, D. de Solla (1963). *Little Science, Big Science*. New York: Columbia University Press, 1963

Quintas, P. and K. Guy (1995). "Collaborative, Pre-competitive R&D and the Firm", *Research Policy,* 24, 325 - 348

Raan, A.F.J. van (2000). "The Pandora's Box of Citation Analysis: Measuring Scientific Excellence, the Last Evil?" in: Cronin, B.; Barskt Atkins, H. (eds.): *The Web of Knowledge. A Festschrift in Honor of Eugene Garfield.* Ch. 15, p. 301-319. Medford (New Jersey): ASIS Monograph Series

Raan, A. Van (1999). "Advanced Bibliometric Methods for the Evaluation of Universities". *Scientometrics*, Vol. 45, No. 3, 417 - 423

Raan, A. Van (1998). "Assessment of Social Sciences: The use of advanced bibliometric methods as a necessary complement of peer review". *Research Evaluation*, Vol. 7, No. 1, 2 - 6

Removille, J. and B. Clarysse (1999). "Intra-European Scientific Co-operation: Measuring policy impact." *Research Evaluation*, August vol. 8, no 2, 99-109

Reka, A. and A Barabasi (2001). "Topology of Evolving Networks: Local events and universality". *Physical Review Letters, 85(24):5234--5237, 11 Dec. 2000*

Rogers, L. A. and J. Anderson (1993). "A New Approach to Defining a Multidisciplinary Field of Science: The Case of Cardiovascular Biology". *Scientometrics*, Vol. 28 No. 1, 61 - 77

Roosendaal, H. and P. A. Geurts (1999). Scientific Communication and its Relevance to Research Policy, *Scientometrics*, Vol. 44, No. 3, 507 - 519

Rosenberg, N. (1990). "Why do Firms do Basic Research (with their own money)?", *Research Policy*, Vol. 19, 165 - 174

Rosenberg, N. (1991). "Critical Issues in Science Policy Research", *Science and Public Policy*, Vol. 18 No. 6, 335 - 346

Sabatino, R. and J. M. de Arce (2000). "Implementation of the Pan-European Academic Research Network: TEN 155", Technical Report on Quantum Project - Quality Network Technology for User-Oriented Multi-media), http://www.dante.net/quantum/, 2000. Document available at http://www.terena.nl/tnnc/4B/4B1/4B1.pdf, in June 2002.

Salvador, M. R. and R. E López-Martínez (2000). "Cognitive Structure of Research: Scientometric Mapping in Sintered Materials". *Research Evaluation*, Vol. 9, No. 3, 189 - 200

Schwartz, L., Glotzer, S., Hall, D. Roosen, A. and J. Warren (1996). "Materials Science in the Information Age", *Technology in Society*, Vol. 18, No. 2, 151 - 164

SearchEngine ShowDown (2002). Statistics for the Size of the Web. http://www.searchengineshowdown.com/stats/sizeest.shtml, March 2002

Shaw, D.F. and H. Moore, Eds. (1996), *Electronic publishing in science*, ICSU Press & UNESCO, 1996

Smarr, Larry (1999). "From supercomputing to the GRID". *Internet 2 Meeting*, Seattle, WA, October 10, 1999

Spasser, M.A. (1997). "Mapping the Terrain of Pharmacy: Co-classification Analysis of the International Pharmaceutical Abstracts Database.", *Scientometrics*, Vol. 39, No. 1, 77 - 97

Steinmueller, W. (1992). "The Economics of Production and Distribution of User-Specific Information via Digital Networks", in C. Antonelli (ed.). *The Economics of Information Networks,* Amsterdam, North-Holland: Elsevier

Steinmueller, W. (1998). "Does Information and Communication Technology facilitate "Codification" of Knowledge?", Working Paper, SPRU, University of Sussex

Steinmueller, W. (1999). "Networked Knowledge and Knowledge-Based Economies", prepared for Telematica Instituut, 'Telematics and the Economy of Information Societies', Delft, Netherlands, 5-6 February

Stump, D. (1996). "From Epistemology and Metaphysics to Concrete Connections", in P. Galison, and D. J. Stump (eds),). *The Disunity of Science. Boundaries, Contexts, and Power*, Stanford, CA, Stanford University Press, pages 255 - 286

Stump, D. (1996). "Afterword: New Directions in the Philosophy of Science Studies", in P. Galison and D. J. Stump (eds), *The Disunity of Science. Boundaries, Contexts, and Power*, Stanford, CA, Stanford University Press, pages 443 - 450

Swanson, D. R. (1968). Scientific Journals and Information Services of the Future. *American Psychologist* Vol. 21, 1005 - 1010

Teece, D (1986), "Profiting from Technological Innovation", *Research Policy*, Vol. 15, 285 - 305

Tijssen, R. and J. C. Korevaar (1997). "Unravelling the Cognitive and Inter-organisational Structure of Public/Private R&D Networks: A Case-Study of Catalysis Research in the Netherlands", *Research Policy,* Vol.25, 1277 - 1293

UCAID (2002). Internet 2. Available at http://www.internet2.edu/, in June 2002

Van Raan, A. (1998). "Assessment of Social Sciences: the Use of Advanced Bibliometric Methods as a Necessary Complement of Peer Review", *Research Evaluation*, Vol. 7, No. 1, 2 - 6
Van Raan, A. (1999). "Advanced Bibliometric Methods for the Evaluation of Universities", *Scientometrics*, Vol. 45, No. 3, 417 - 423

Von Hippel, E. (1976). "The Dominant Role of Users in the Scientific Instrument Innovation Process", *Research Policy*, Vol. 5, 212 - 239

Von Hippel, E. and M. J. Tyre (1995). "How Learning by Doing is Done: Problem Identification in Novel Process Equipment", *Research Policy*, Vol. 24, 1 - 12

Walsh, J. P. and T. Bayma (1996). "Computer Networks and Scientific Work", *Social Studies of Science*, Vol. 26, 661 - 703

Wassserman, S. and K. Faust (1994). *Social Network Analysis. Methods and Applications*. Cambridge, Cambridge University Press.

Watts, D. J. and Strogatz, S. H. (1998). "Collective Dynamics of 'Small-World' networks". *Nature*, Vol. 393, June, 440 - 442

Watts, D. J. (1999). *Small-Worlds . The Dynamics of Networks from Order to Randomness*. Princeton University Press.

Wellman, B. (1996). "For a Social Network Analysis of Computer Networks: A Sociological Perspective on Collaborative Work and Virtual Community*", in Proceedings* of SIGCPR/SIGMIS '96, Denver Colorado USA, ACM.

Widhalm, C., Topolnik, M., Kopcsa, A., Schiebel, E. and M. Weber (2001). "Evaluating Patterns of Co-operation: Application of a Bibliometric Visualisation Tool to the Fourth Framework Programme and the Transport Research Programme". *Research Evaluation*, August 129 - 140

Whitley, R. (2000), *The Intellectual and Social Organisation of the Sciences*, Oxford, Oxford University Press

Whitley, R. (1984), "The Fragmented State of Management Studies: Reasons and Consequences", *Journal of Management Studies*, Vol. 21, 331 - 348

Whitley, Richard D. (1970). "The Formal Communication System of Science: a Study of the Organisation of British Social Science Journals", in P. Halmos (ed.) *Sociological Review Monograph, No. 16* London: Routledge and Kegan Paul.

Woolgar, S. (1992). "Some Remarks about Positionism. A Reply to Collins and Yearley", in A. Pickering (ed.), *Science as Practice and Culture*, Chicago:University of Chicago Press, pages 327 - 342

Wouters, P. (1999). "Beyond the Holy Grail: from Citation Theory to Indicator Analysis", *Scientometrics*, Vol. 44, No. 3, 561 - 580

Ziman J. (1984), *An Introduction to Science Studies: The Philosophical and Social Aspects of Science and Technology*, Cambridge, Cambridge University Press

Ziman, J. M. (1980). *The Proliferation of Scientific Literature: A Natural Process*. AAAS, 1980

Zuckerman, H and R. K. Merton (1971). "Patterns of Evaluation in Science: Institutionalisation, Structure and Functions of the Referee System", *Minerva* Vol. 9, no. 1, 66 - 100

Zulueta, M. A and M. Bordons (1999). "A Global Approach to the Study of Teams in Multidisciplinary Research Areas through Bibliometric Indicators". *Research Evaluation*, Vol. 8, No. 2, 111 - 118

8. Annexes

ANNEX I
Evolution of Publication in Automatic Speech Recognition

Importance of more Informal communication - Evolution of number of articles indexed in Technical Proceeding of Conferences, based on SCI bibliographic database.

Table I.1 - Bibliometric data on "Automatic Speech Recognition"

	Science Citation Index		Technical Proceedings	
Year	Nr. of Publications	Cumul.Nr.Publications	Nr.of Publications	Cumul.Nr.Publications
1981	16	16	-	-
1982	29	45	35	35
1983	46	91	36	71
1984	34	125	16	87
1985	41	166	47	134
1986	26	192	92	226
1987	36	228	81	307
1988	23	251	67	374
1989	30	281	47	421
1990	45	326	83	504
1991	171	497	130	634
1992	166	663	115	749
1993	162	825	87	836
1994	218	1 043	80	916
1995	272	1 315	224	1 140
1996	274	1 589	117	1 257
1997	285	1 874	605	1 862
1998	357	2 231	427	2 289
1999	393	2 624	320	2 609
2000	419	3 043	333	2 942
Total	**3 043**		**2 942**	

Source: Data analyses from database records on SCI - Web of Science (http://wos.mimas.ac.uk/), January 2001

ANNEX II

Co-authorship Networks, based on Rabiner's seminal article Bibliographic-coupled publications

Cliques identified based on Co-Authorship Network of Bibliographic-coupled publications with Rabiner's 1989 article in journal *Proceedings of the IEEE*.

The strength of association is defined as greater than or equal to 2:

In the period - 1981 - 1988

1 clique found.

1: RABINER_LR LEVINSON_SE SONDHI_MM JUANG_BH
BELL TEL LABS INC,MURRAY HILL,NJ 07974
Later AT&T BELL LABS,MURRAY HILL,NJ 07974

In the period 1989 - 2000

11 Cliques found

1: RUNKLE_P CARIN_L YODER_TJ BUCARO_JA
Duke Univ, Dept Elect & Comp Engn, Durham, NC 27708 USA
Research Lab, Washington - USA
SFA Inc, Landover, MD 20785 USA

2: RUNKLE_P CARIN_L COUCHMAN_L
Duke Univ, Dept Elect & Comp Engn, Durham, NC 27708 USA
Research Lab, Washington - USA

3: CARIN_L BHARADWAJ_PK RUNKLE_PR
Duke Univ, Dept Elect & Comp Engn, Durham, NC 27708 USA

4: JUANG_BH RABINER_LR LEE_CH
AT&T BELL LABS,MURRAY HILL,NJ 07974

5: DENG_L LENNIG_M MERMELSTEIN_P
INRS TELECOMMUN,3 PL COMMERCE,MONTREAL H3E 1H6,QUEBEC,CANADA;
BELL NO RES,MONTREAL,QUEBEC,CANADA

6: COAST_DA CANO_GG BRILLER_SA
ALLEGHENY SINGER RES INST,PITTSBURGH,PA 15212

7: MAKHOUL_J SCHWARTZ_R BAZZI_I
BBN Syst & Technol Corp, GTE Internetworking, 70 Fawcett St,;Cambridge, MA 02138 USA;
BBN Syst & Technol Corp, GTE Internetworking, Cambridge, MA 02138 USA

8: STOLCKE_A SHRIBERG_E HAKKANI-TUR_D TUR_G
SRI Int, Speech Technol & Res Lab, 333 Ravenswood Ave, Menlo;Park, CA 94025 USA;
SRI Int, Speech Technol & Res Lab, Menlo Park, CA 94025 USA;
Bilkent Univ, Dept Comp Engn, TR-06533 Ankara, Turkey

9: KWONG_S MAN_KF TANG_KS
City Univ Hong Kong, Dept Comp Sci, Tatchee Ave, Kowloon, Hong;Kong, Peoples R China;
City Univ Hong Kong, Dept Comp Sci, Kowloon, Hong Kong, Peoples R China;
City Univ Hong Kong, Dept Elect Engn, Kowloon, Hong Kong, Peoples R China

10: SHAKIBA_MH JOHNS_DA MARTIN_KW
Gennum Corp, Burlington, ON, Canada;
Univ Toronto, Dept Elect & Comp Engn, Toronto, ON M5S 3G4, Canada

11: VENKATARAMANAN_L KUC_R SIGWORTH_FJ
Yale Univ, Dept Elect Engn, New Haven, CT 06520 USA;
Yale Univ, Dept Elect Engn, New Haven, CT 06520 USA;
Yale Univ, Sch Med, Dept Cellular & Mol Physiol, New Haven, CT 06520 USA

ANNEX III
Collaboration Groups detected by Co-authorship analysis of 500 "citing documents" to Rabiner seminal article in 1989

Twenty-five different Collaboration groups, identified by co-authorship analysis, were detected.
25 cliques found.
1: SHUE_L ANDERSON_BDO DE BRUYNE_F
Nanyang Technol Univ, Ctr Signal Proc, Singapore 2263,;Singapore;
Nanyang Technol Univ, Ctr Signal Proc, Singapore 2263, Singapore;
Australian Natl Univ, Dept Syst Engn, Canberra, ACT 2601, Australia;
Siemens Brussels, Adv Proc Control Grp, Huizingen, Belgium

2: ANTON-HARO_C FONOLLOSA_JAR FONOLLOSA_JR
Univ Politecn Catalunya, Dept Signal Theory & Commun,;Barcelona, Spain;

3: WONG_JC MCDONALD_KA PALAZOGLU_A
Univ Calif Davis, Dept Chem Engn & Mat Sci, Davis, CA 95616 USA;

4: OTTERPOHL_JR EMMERT-STREIB_F PAWELZIK_K
Univ Bremen, Inst Theoret Neurophys, Ctr Cognit Sci, Kufsteiner;Str, D-28334 Bremen, Germany;
Univ Bremen, Inst Theoret Neurophys, Ctr Cognit Sci, D-28334 Bremen, Germany;
Univ Bremen, Ctr Cognit Sci, D-28334 Bremen, Germany

5: ABELES_M SEIDEMANN_E MEILIJSON_I BERGMAN_H VAADIA_E
HEBREW UNIV JERUSALEM,SCH MED,POB 12272,IL-91120;JERUSALEM,ISRAEL
TEL AVIV UNIV,RAYMOND & BEVERLY SACKLER FAC EXACT SCI,SCH MATH;SCI,IL-69978 RAMAT AVIV,ISRAEL
TEL AVIV UNIV,SCH MED,IL-69978 RAMAT AVIV,ISRAEL
HEBREW UNIV JERUSALEM,CTR NEURAL COMPUTAT,IL-91120 JERUSALEM,ISRAEL
TEL AVIV UNIV,SCH MED,IL-69978 TEL AVIV,ISRAEL

6: ABELES_M GAT_I TISHBY_N
HEBREW UNIV JERUSALEM,HADASSAH MED SCH,IL-91904 JERUSALEM,ISRAEL
HEBREW UNIV JERUSALEM,INST COMP SCI,IL-91904 JERUSALEM,ISRAEL
HEBREW UNIV JERUSALEM,CTR NEURAL COMPUTAT,IL-91904 JERUSALEM,ISRAEL

7: RUNKLE_P CARIN_L YODER_TJ BUCARO_JA
Duke Univ, Dept Elect & Comp Engn, Durham, NC 27708 USA;Duke Univ, Dept Elect & Comp Engn, Durham, NC 27708 USA
USN, Res Lab, Washington, DC 20375 USA
SFA Inc, Landover, MD 20785 USA

8: RUNKLE_P COUCHMAN_L CARIN_L
Duke Univ, Sch Engn, Dept Elect & Comp Engn, Box 90291, Durham,;NC 27708 USA;
Duke Univ, Sch Engn, Dept Elect & Comp Engn, Durham, NC 27708 USA
Duke Univ, Sch Engn, Dept Elect & Comp Engn, Durham, NC 27708 USA

9: CARIN_L RUNKLE_PR BHARADWAJ_PK
Duke Univ, Dept Elect & Comp Engn, Durham, NC 27708 USA

10: LEE_CH JUANG_BH CHOU_W
AT&T BELL LABS,SPEECH RES DEPT,MURRAY HILL,NJ 07974

11: DEHGHAN_M FAEZ_K AHMADI_M SHRIDHAR_M
Univ Windsor, Dept Elect & Comp Engn, Windsor, ON N9B 3P4,;Canada;
Univ Windsor, Dept Elect & Comp Engn, Windsor, ON N9B 3P4, Canada;
Amirkabir Univ Technol, Dept Elect Engn, Tehran, Iran;
Univ Michigan, Dept Elect & Comp Engn, Dearborn, MI 48128 USA

12: KSCHISCHANG_FR SOROKINE_V PASUPATHY_S
Qualcomm Inc, San Diego, CA 92121 USA;
Univ Toronto, Dept Elect & Comp Engn, Toronto, ON M5S 3G4, Canada

13: SCHWARTZ_R BAZZI_I MAKHOUL_J
GTE Internetworking, BBN Technol, Cambridge, MA 02138 USA

14: MULLER_S EICKELER_S RIGOLL_G
Univ Duisburg Gesamthsch, Fac Elect Engn, Dept Comp Sci, D-4100;Duisburg, Germany;

15: SHUE_L DEY_S ANDERSON_BDO

16: HAZOUT_S CAMPROUX_AC TUFFERY_P BOISVIEUX_JF
Univ Paris 07, INSERM, U155, Equipe Bioinformat Mol, Case;7113,2 Pl Jussieu, F-75251 Paris 05, France;
CHU Pitie Salpetriere, Dept Biomath, F-75013 Paris, France

17: QIN_F AUERBACH_A SACHS_F
SUNY BUFFALO,DEPT BIOPHYS SCI,BUFFALO,NY 14214

18: LECUN_Y HAFFNER_P BOTTOU_L BENGIO_Y
AT&T Shannon Lab, 100 Schulz Dr, Red Bank, NJ 07701 USA;
AT&T Shannon Lab, Red Bank, NJ 07701 USA

19: KIM_HJ KIM_SK LEE_JK KIM_KH
Inje Univ, Dept Comp Sci, Kimhae 621749, South Korea;
Andong Natl Univ, Dept Comp Engn, Andong 760600, South Korea;
Kyungpook Natl Univ, Dept Comp Engn, Taegu 702701, South Korea

20: MIN_BW YOON_HS SOH_J OHASHI_T EJIMA_T
ETRI, CSTL, Image Proc Dept, Yuseong Gu, 161 Kajong Dong,;Taejon 305350, South Korea;
ETRI, CSTL, Image Proc Dept, Yuseong Gu, Taejon 305350, South Korea;
Kyushu Inst Technol, Dept Artificial Intelligence, Iizuka, Fukuoka 820, Japan

21: MIAN_IS MOSER_MJ HOLLEY_WR CHATTERJEE_A
Univ Calif Berkeley, Lawrence Berkeley Lab, Div Life Sci, Mail;Stop 29-100,1 Cyclotron Rd, Berkeley, CA 94720 USA;
Univ Calif Berkeley, Lawrence Berkeley Lab, Div Life Sci, Berkeley, CA 94720 USA

22: SHARMA_R PAVLOVIC_VI HUANG_TS
Penn State Univ, Dept Comp Sci & Engn, University Pk, PA 16802 USA;
Univ Illinois, Beckman Inst, Urbana, IL 61801 USA;
Univ Illinois, Dept Elect & Comp Engn, Urbana, IL 61801 USA

23: GRUNDY_WN BAILEY_TL ELKAN_CP BAKER_ME
UNIV CALIF SAN DIEGO,DEPT COMP SCI & ENGN,LA JOLLA,CA 92093;
UNIV CALIF SAN DIEGO,DEPT MED,LA JOLLA,CA 92093;SAN DIEGO SUPERCOMP CTR,SAN DIEGO,CA 92186

24: PEINADO_AM RUBIO_AJ SEGURA_JC GARCIA_P
UNIV GRANADA,FAC CIENCIAS,DIPARTIMENTO ELECT & TECHNOL COMP,E-;18071 GRANADA,SPAIN

25: PEINADO_AM RUBIO_AJ SANCHEZ_VE
UNIV GRANADA,FAC CIENCIAS,DIPARTIMENTO ELECT & TECHNOL COMP,E-;18071 GRANADA,SPAIN

ANNEX IV
Co-authorship networks - Language research - Continuum Approach (1981 - 1992)

Co-authorship Network (1981 - 1990)
Collaboration groups (only 1 collaboration)
58 cliques found.

1: CHANG_SK JUNGERT_E LEVIALDI_S TORTORA_G ICHIKAWA_T
2: TORTORA_G CRIMI_C GUERCIO_A PACINI_G TUCCI_M
3: MIWA_T NAKAGAWA_K FUJIMOTO_F IKUNO_T
4: HOVY_EH MCDONALD_DD YOUNG_SR
5: WEISCHEDEL_R CARBONELL_J GROSZ_B LEHNERT_W MARCUS_M PERRAULT_R WILENSKY_R
6: REFENES_AN EBERBACH_E MCCABE_SC TRELEAVEN_PC
7: OZSOYOGLU_G MATOS_V OZSOYOGLU_ZM
8: BOUMA_G KONIG_E USZKOREIT_H
9: KIKUCHI_DT MIRANDA_RF THYSELL_PA
10: CROOKES_D MORROW_PJ MILLIGAN_P KILPATRICK_PL SCOTT_NS
11: CROOKES_D MILLIGAN_P PERROTT_RH PURDY_WRM
12: CROOKES_D MORROW_PJ MCPARLAND_PJ
13: TSUNO_K KOKUBO_Y YANAGAWA_H
14: HIRSCHBERG_J BALLARD_BW HINDLE_D
15: BALLARD_BW LUSTH_JC TINKHAM_NL
16: DORDA_W HAIDL_B SACHS_P
17: NAKANO_K SAKAGUCHI_Y ISOTANI_R OHMORI_T
18: CARAMAZZA_A BASILI_AG KOLLER_JJ BERNDT_RS
19: VANLIESHOUT_PHHM RENIER_W ELING_P SLIS_I DEBOT_K
20: RENIER_W ELING_P SLIS_I DEBOT_K VANLIESHOUT_P
21: JAREMA_G KADZIELAWA_D WAITE_J
22: LOMAS_J PICKARD_L MOHIDE_A
23: DELOCHE_G SERON_X SCIUS_G SEGUI_J
24: DE_SJ PAN_SS WHINSTON_A
25: TEST_JA MYSZEWSKI_M SWIFT_RC
26: TAKEUCHI_I OKUNO_H OHSATO_N
27: LEGUERNIC_P BENVENISTE_A BOURNAI_P GAUTIER_T
28: ALTENMULLER_E JUNG_R LANDWEHRMEYER_B
29: METZLER_DP NOREAULT_T HAAS_DF COSIC_C
30: PIGUET_C DIJKSTRA_E BERWEILER_G
31: BROWN_JW BARTLETT_EJ WOLF_AP RUSSELL_J BRODIE_J
32: PROTSENKO_VS RODIMIN_SP STAVROVSKII_AB KHIZHNYAK_AA YURCHISHIN_VV
33: SAGER_N PETRICK_SR BRINER_LL KITTREDGE_R BORKO_H
34: BREGUET_P GRIZE_F STROHMEIER_A
35: JAY_JM VANZANDT_RN HANSON_PG
36: ZIEGLER_E SCHOMBURG_G WEIMANN_B WRONKA_I HAUSIG_U
37: DANLY_M COOPER_WE SHAPIRO_B
38: ROTHENBERG_A SZIRTES_J JURGENS_R
39: BOHM_M NICOLAE_GC HOHNE_KH
40: TRACY_P CHIDLAW_R GONSALVES_R
41: NINOMIYA_H NAKAHARA_D IKEDA_T
42: SEITZ_MR WEBER_BA JACOBSON_JT MOREHOUSE_R
43: MIYAGUCHI_S OHTA_K IWATA_M
44: CLEMENTE_F CESARELLI_M BRACALE_M
45: HOWELL_SV BAVUSO_SJ HALEY_PJ
46: ALEXIN_Z GYIMOTHY_T HORVATH_T FABRICZ_K
47: STEKOLNIKOV_VV TOLMACHEVA_MY SAVENKOV_VM
48: JAYANT_NS LAWRENCE_VB PREZAS_DP

49: CLAUSEN_M LIETZENMAYER_R WELLER_R OCHSENKUHN_R LEUTLOFF_UC WURZ_C HENZE_E ADAM_WE
50: SATO_J KUWAMURA_Y OHNO_C
51: HEERING_J KLINT_P REKERS_J
52: BARNETT_J KNIGHT_K MANI_I RICH_E
53: STEPHANOPOULOS_G HENNING_G LEONE_H
54: BHARATI_A SANGAL_R CHAITANYA_V
55: YANNAKOUDAKIS_EJ TSOMOKOS_I HUTTON_PJ
56: GALLOWAY_A BIRKBY_WH KAHANA_T FULGINITI_L
57: ILLES_J CUTILLO_BA BRESSLER_SL GEVINS_AS
58: FURUYAMA_T OKAMOTO_T HOSOYA_R

Collaboration groups (more than 1 collaboration)
1 cliques found.

1: RENIER_W ELING_P SLIS_I DEBOT_K
UNIV NIJMEGEN,DEPT EXPTL PSYCHOL,6500 HE NIJMEGEN,NETHERLANDS

Co-authorship Network - Language Research (1981 - 1991)

Collaboration groups (more than 1 collaboration)
3 cliques found.

1: CROOKES_D MORROW_PJ MCPARLAND_PJ
QUEENS UNIV BELFAST,DEPT COMP SCI,BELFAST BT7 1NN,ANTRIM,NORTH;IRELAND

2: RENIER_W ELING_P SLIS_I DEBOT_K
UNIV NIJMEGEN,DEPT EXPTL PSYCHOL,6500 HE NIJMEGEN,NETHERLANDS

3: NUMAZAKI_H TAMURA_N TANAKA_H
YOKOHAMA NATL UNIV,YOKOHAMA,JAPAN;TOKYO INST TECHNOL,MEGURO KU,TOKYO 152,JAPAN

Co-authorship Network (1981 - 1992)
Collaboration groups (more than 1 collaboration)
14 cliques found.
1: WISE_R FRISTON_K FRACKOWIAK_R
HAMMERSMITH HOSP,MRC,CYCLOTRON UNIT,LONDON W12 0HS,ENGLAND
UNIV CAMBRIDGE,MRC,APPL PSYCHOL UNIT,CAMBRIDGE CB2 1TN,ENGLAND

2: WISE_R CHOLLET_F FRACKOWIAK_R
HAMMERSMITH HOSP,MRC,CYCLOTRON UNIT,LONDON W12 0HS,ENGLAND
UNIV CAMBRIDGE,MRC,APPL PSYCHOL UNIT,CAMBRIDGE CB2 1TN,ENGLAND

3: WISE_R FRACKOWIAK_R RAMSAY_S
HAMMERSMITH HOSP,MRC,CYCLOTRON UNIT,LONDON W12 0HS,ENGLAND
UNIV CAMBRIDGE,MRC,APPL PSYCHOL UNIT,CAMBRIDGE CB2 1TN,ENGLAND
NUCL MED & ULTRASOUND ASSOCIATES,ASHLEY CTR,WESTMEAD,NSW,AUSTRALIA

4: WISE_R HOWARD_D PATTERSON_K
HAMMERSMITH HOSP,MRC,CYCLOTRON UNIT,LONDON W12 0HS,ENGLAND
Inst Neurol, Wellcome Dept Cognit Neurol, London WC1N 3BG, England
UNIV LONDON BIRKBECK COLL,DEPT PSYCHOL,LONDON WC1E 7HX,ENGLAND

5: CROOKES_D MORROW_PJ MCPARLAND_PJ
QUEENS UNIV BELFAST,DEPT COMP SCI,BELFAST BT7 1NN,ANTRIM,NORTH;IRELAND

6: RENIER_W ELING_P SLIS_I DEBOT_K
UNIV NIJMEGEN,DEPT EXPTL PSYCHOL,6500 HE NIJMEGEN,NETHERLANDS

7: IKEDA_T HERATH_S ISHIZAKI_S ANZAI_Y AISO_H
SW TEXAS STATE UNIV,SAN MARCOS,TX 78666

8: HEERING_J KLINT_P REKERS_J
CENT WISKUNDE INFORMAT,POB 4079,1009 AB AMSTERDAM,NETHERLANDS;
UNIV AMSTERDAM,AMSTERDAM,NETHERLANDS

9: SUMITA_K UKITA_T AMANO_S
TOSHIBA CO LTD,CTR RES & DEV,KAWASAKI 210,JAPAN

10: BROWN_HK HAZELTON_TR SILBIGER_ML
UNIV S FLORIDA,COLL MED,DEPT ANAT,TAMPA,FL 33612;
UNIV S FLORIDA,COLL MED,DEPT RADIOL,TAMPA,FL 33612

11: BECHTEREVA_NP ABDULLAEV_YG MEDVEDEV_SV
INST EXPTL MED,DEPT HUMAN NEUROPHYSIOL,LENINGRAD,USSR

12: NUMAZAKI_H TAMURA_N TANAKA_H
YOKOHAMA NATL UNIV,YOKOHAMA,JAPAN;TOKYO INST TECHNOL,MEGURO KU,TOKYO 152,JAPAN

13: QU_DL MURAKI_M HAYAKAWA_T
TOKYO INST TECHNOL,DEPT IND ENGN & MANAGEMENT,TOKYO 152,JAPAN

14: LANE_TA ANDERSON_KC GOODNOUGH_LT SILBERSTEIN_LE
UNIV CALIF SAN DIEGO,SCH MED,DEPT PATHOL 0612,9500 GILMAN DR,LA;JOLLA,CA 92093;
HARVARD UNIV,SCH MED,DANA FARBER CANC INST,BOSTON,MA 02115;
UNIV HOSP CLEVELAND,DEPT MED,CLEVELAND,OH 44106
HOSP UNIV PENN,BLOOD BANK,PHILADELPHIA,PA 19104

ANNEX V
Collaboration groups identified by simultaneous publication in both bibliometric data-sets (Speech and Language)
7 cliques found.

1: PLACE_JF TRUCHAUD_A OZAWA_K PARDUE_H SCHNIPELSKY_P
DAKO AS, MA, PROD VEJ 42, DK-2600 GLOSTRUP, DENMARK

2: PEREIRA_F MOHRI_M RILEY_M
AT&T Bell Labs, Res, Florham Park, NJ 07932 USA.
AT&T Bell Labs, Res, 180 Pk Ave, Florham Park, NJ 07932 USA.
AT&T Labs Res, Room E141, 180 Pk, Florham Park, NJ 07932 USA.

3: RICCARDI_G GORIN_AL WRIGHT_JH
AT&T Bell Labs, Res, Florham Park, NJ USA.;AT&T Bell Labs, Res, Florham Park, NJ USA.

4: OVIATT_S LEVOW_GA MACEACHERN_M
Oregon Grad Inst Sci & Technol, Dept Comp Sci, Ctr Human Comp;Commun, POB 91000, Portland, OR 97291 USA.;
Oregon Grad Inst Sci & Technol, Dept Comp Sci, Ctr Human Comp Commun, Portland, OR 97291 USA.;
Univ Pittsburgh, Dept Linguist, Pittsburgh, PA 15260 USA.;
MIT, Artificial Intelligence Lab, Cambridge, MA 02139 USA.

5: SHIFFMAN_S DETMER_WM LANE_CD FAGAN_LM
STANFORD UNIV, MED INFORMAT SECT, MED SCH OFF BLDG X215,;STANFORD, CA 94305.

6: KESSLER_DK LOEB_GE OSBERGER_MJ
UNIV CALIF SAN FRANCISCO, SCH MED, DEPT OTOLARYNGOL, 400;PARNASSUS AVE, ROOM A-701, SAN FRANCISCO, CA 94143.

7: CHOPE_M METZLUTZ_MN WIOLAND_N RUMBACH_L KURTZ_D
HOP UNIV STRASBOURG, NEUROL CLIN, SERV EXPLORAT FONCT SYST;NERVEUX, F-67091 STRASBOURG, FRANCE.;
HOP UNIV STRASBOURG, NEUROL CLIN, SERV NEUROPSYCHOL & REEDUCAT LANGAGE, F-67091 STRASBOURG, FRANCE.;
CHU BESANCON, HOP J MINJOZ, NEUROL SERV, F-25030 BESANCON, FRANCE.

ANNEX VI
Full List of Collaboration Groups in Speech Research

As given by co-authorship networks (1999 - 2001), Extended Broad Search

39 cliques found.

1: VAN SON_RJJH POLS_LCW
Univ Amsterdam, Inst Phonet Sci, IFOTT, NL-1016 CG Amsterdam, Netherlands

2: DE LA HIGUERA_C CASACUBERTA_F
Departamento de Sistemas Informáticos y Computación, Universidad Politécnica, Valencia, Spain

3: DE LA TORRE_A
Univ Granada, Fac Ciencias, Dept Elect & Tecnol Comp, E-18071;Granada, Spain

4: DE VETH_J CRANEN_B BOVES_L
Univ Nijmegen, Dept Language & Speech, A2RT,POB 9103, NL-6500;HD Nijmegen, Netherlands

5: DE CHEVEIGNE_A KAWAHARA_H
Lab Linguist Formelle, Case 7003,2 Pl Jussieu, F-75251 Paris,;France;Univ Paris 07, CNRS, F-75251 Paris, France;ATR Human Informat Proc Res Labs, Kyoto 6190288, Japan;Wakayama Univ, Fac Syst Engn, Design Informat Sci Dept, Media Design Informat Grp, Wakayama 6408510, Japan

6: BOVES_L CUCCHIARINI_C STRIK_H
Univ Nijmegen, Dept Language & Speech, POB 9103, NL-6500 HD;Nijmegen, Netherlands;Univ Nijmegen, Dept Language & Speech, NL-6500 HD Nijmegen, Netherlands

7: LEE_CH LI_Q JUANG_BH
AT&T Bell Labs, Lucent Technol, Murray Hill, NJ 07974 USA
Lucent Technol, Bell Labs, Dialogue Syst Res Dept, Room 2C-572,;600 Mt Ave, Murray Hill, NJ 07974 USA
AT&T Bell Labs, Lucent Technol, Murray Hill, NJ 07974 USA

8: HUO_Q JIANG_H HIROSE_K
Univ Hong Kong, Dept Comp Sci & Informat Syst, Hong Kong, Hong Kong, Peoples R China
Bell Labs, Lucent TEchnol, Dialogue Syst Res Dept, 600 Mt Ave,;Murray Hill, NJ 07974 USA;Univ Tokyo, Dept Informat & Commun Engn, Tokyo 1130033, Japan
Univ Tokyo, Dept Frontier Informat, Tokyo 1130033, Japan

9: GULMEZOGLU_MB DZHAFAROV_V BARKANA_A
Osmangazi Univ, Elect & Elect Engn Dept, Eskisehir, Turkey

10: MORGAN_DR BENESTY_J CHO_JH
Bell Labs, Lucent Technol, 600 Mt Ave, Murray Hill, NJ 07974;USA

11: BENESTY_J GAY_SL GANSLER_T
Bell Labs, Lucent Technol, Murray Hill, NJ 07974 USA
Univ Lund, Dept Appl Elect, S-22100 Lund, Sweden
Univ Lund, Dept Appl Elect, S-22100 Lund, Sweden

12: KIM_HK CHOI_SH LEE_HS
AT&T Labs Res, Florham Park, NJ 07932 USA

13: ESCH_O VANSONNENBERG_E GOODACRE_BW OZKAN_OS WITTICH_GR
20TH AND NORTHAMPTON STS, EASTON, PA 18042 USA (Editor Address)

14: NGUYEN_P KUHN_R JUNQUA_JC NIEDZIELSKI_N

Panason Technol, Speech Technol Lab, Suite 202, 3888 State St,;Santa Barbara, CA 93105 USA.;Panason Technol, Speech Technol Lab, Santa Barbara, CA 93105 USA

15: BAUMGARTNER_WD ADUNKA_O GSTOETTNER_W HAMZAVI_J
Univ Klin Hals Nasen & Ohrenkrankheiten, A-1097 Vienna, Austria.

16: STROUSE_A WILSON_RH BRUSH_N
VA Med Ctr, Mt Home, TN 37684 USA.;VA Med Ctr, Mt Home, TN 37684 USA.;
E Tennessee State Univ, Dept Surg, Johnson City, TN 37614 USA.;
E Tennessee State Univ, Dept Commun Disorders, Johnson City, TN 37614 USA.

17: DIGALAKIS_V NEUMEYER_L FRANCO_H
SRI Int, Speech Technol & Res Lab, 333 Ravenswood Ave, Menlo;Park, CA 94025 USA

18: KHUDANPUR_S SARACLAR_M NOCK_H

Univ Cambridge, Dept Engn, Cambridge CB2 1PZ, England.
Johns Hopkins Univ, Ctr Language & Speech Proc, Baltimore, MD;21218 USA.
Johns Hopkins Univ, Ctr Language & Speech Proc, Baltimore, MD 21218 USA

19: KHUDANPUR_S BYRNE_W MCDONOUGH_J
Univ Cambridge, Dept Engn, Cambridge CB2 1PZ, England.
AT&T Labs Res, Florham Park, NJ 07932 USA
Univ Cambridge, Dept Engn, Cambridge CB2 1PZ, England.

20: LOIZOU_PC DORMAN_M POROY_O
Univ Arkansas, Dept Appl Sci, Little Rock, AR 72204 USA.(then moved to U Texas?)
ARIZONA STATE UNIV, DEPT SPEECH & HEARING SCI, TEMPE, AZ 85287
Univ Texas, Dept Elect Engn, Richardson, TX 75083 USA

21: CLARK_GM GALVIN_KL BLAMEY_PJ COWAN_RSC OERLEMANS_M
Univ Melbourne, Dept Otolaryngol, RVEEH, E Melbourne, Vic 3002, Australia.
CRC Cochlear Implant Speech & Hearing Res, 384-388 Albert St, E;Melbourne, Vic 3002, Australia
CRC Cochlear Implant Speech & Hearing Res, E Melbourne, Vic 3002, Australia
Univ Melbourne, Dept Otolaryngol, RVEEH, E Melbourne, Vic 3002, Australia.
Bion Ear Inst, E Melbourne, Vic 3002, Australia.

22: NEUMAN_AC LEVITT_H MACKERSIE_C

CUNY, CTR RES SPEECH & HEARING SCI, 33 W 42ND ST, NEW YORK, NY 10036. (Levitt 1991)
San Diego State Univ, Dept Communicat Disorders, 5500 Campanile Dr, San Diego, CA 92182 USA.

23: JUNG_HY LEE_TW LEE_SY
Korea Adv Inst Sci & Technol, Brain Sci Res Ctr, Yusung Gu, Taejon 305701, South Korea.
Korea Adv Inst Sci & Technol, Dept Elect Engn, Yusung Gu, Taejon 305701, South Korea.;
Univ Calif San Diego, Inst Neural Computat, La Jolla, CA 92093 USA.

24: WANG_G SKINNER_MW RUBINSTEIN_JT VANNIER_MW
Univ Iowa, Dept Radiol, 200 Hawkins Dr, Iowa City, IA 52242;USA.;
Univ Iowa, Dept Radiol, Iowa City, IA 52242 USA.
Washington Univ, Sch Med, Dept Otolaryngol Head & Neck Surg, St Louis, MO 63110 USA.
Univ Iowa, Dept Surg, Iowa City, IA 52242 USA.

25: KAJITA_S TAKEDA_K ITAKURA_F
Nagoya Univ, Ctr Informat Media Studies, Nagoya, Aichi 4648603, Japan.

26: STRASSMANN_G WALTER_S KOLOTAS_C HEYD_R BALTAS_D SAKAS_G ZAMBOGLOU_N
Stadt Kliniken Offenbach, Strahlenklin, Dept Radiat Med, D-63069 Offenbach, Germany.;
Stadt Kliniken Offenbach, Strahlenklin, Dept Radiat Med, D-63069 Offenbach, Germany.;
Fraunhoferinst, Inst Graph Datenverarbeitung, Darmstadt, Germany.
Natl Tech Univ Athens, Inst Commun & Comp Syst, Athens, Greece.

27: STUDEBAKER_GA SHERBECOE_RL MCDANIEL_DM
Univ Memphis, Memphis Speech & Hearing Ctr, 807 Jefferson Ave,;Memphis, TN 38105 USA
Univ Memphis, Memphis Speech & Hearing Ctr, Memphis, TN 38105 USA.;
Arkansas State Univ, State Univ, AR 72467 USA.

28: MEHTA_A DREYER_KJ THRALL_JH
20TH AND NORTHAMPTON STS, EASTON, PA 18042 USA (Editor's Address)

29: SNYDER_RL VOLLMER_M MOORE_CM REBSCHER_SJ LEAKE_PA BEITEL_RE
Univ Calif San Francisco, Epstein Lab, Box 0526, San Francisco,;CA 94143 USA.;
Univ Calif San Francisco, Epstein Lab, San Francisco, CA 94143 USA.

30: VERSFELD_NJ FESTEN_JM HOUTGAST_T
Univ Amsterdam, Acad Med Ctr, Room D2-330, Meibergdreef 9, NL-;1105 AZ Amsterdam, Netherlands.;Free
Univ Amsterdam Hosp, Dept Otorhinolaryngol, NL-1007 MB Amsterdam, Netherlands.

31: PALLER_KA GRABOWECKY_M BOZIC_VS YAMADA_S
Northwestern Univ, Dept Psychol, Evanston, IL 60208 USA.

32: HE_QH KWONG_S MAN_KF TANG_KS
City Univ Hong Kong, 83 Tat Chee Ave, Kowloon, Hong Kong.;
City Univ Hong Kong, Kowloon, Hong Kong.;
S China Univ Technol, Guangzhou, Peoples R China.

33: MARTIN_GI FURMAN_GI SINDEL_BD
Neonatol, W Covina, CA USA.

34: SCHRAMM_H RUEBER_B KELLNER_A
Philips Res Labs, POB 1980, D-52021 Aachen, Germany;
Philips Res Labs, D-52021 Aachen, Germany

35: COHEN_NL WALTZMAN_SB ROLAND_JT
NYU, SCH MED, DEPT OTOLARYNGOL, 550 1ST AVE, NEW YORK, NY;10016.

36: HODGES_AV BALKANY_TJ GOMEZ-MARIN_O BUTTS_S LEE_D
Box 016960 D-48, Miami, FL 33101 USA.;
Univ Miami, Ear Inst, Dept Otolaryngol Head & Neck Surg, Miami, FL 33152 USA.;
Univ Miami, Ear Inst, Dept Epidemiol & Publ Hlth, Miami, FL 33152 USA.

37: HODGES_AV BUTTS_S LEE_D BIRD_PA
Box 016960 D-48, Miami, FL 33101 USA.;
Univ Miami, Ear Inst, Dept Otolaryngol Head & Neck Surg, Miami, FL 33152 USA.;
Univ Miami, Ear Inst, Dept Epidemiol & Publ Hlth, Miami, FL 33152 USA.

38: PAVLICEK_W MUHM_JR COLLINS_JM ZAVALKOVSKIY_B PETER_BS HINDAL_MD
Mayo Clin Scottsdale, Dept Radiol, 13400 Shea Blvd, Scottsdale,;AZ 85259 USA.;Mayo Clin Scottsdale, Dept Radiol, Scottsdale, AZ 85259 USA.;
Mayo Clin Scottsdale, Dept Clin Engn, Scottsdale, AZ 85259 USA.

39: HANEDA_Y KANEDA_Y KITAWAKI_N
ENGINEERING CENTER 2613-1 ICHINOMOTO-CHO TENRISHI, NARA 632,;JAPAN

Collaboration Groups Speech - CoreAuthors (1999 - 2001)
24 cliques found.

1: SHUE_L ANDERSON_BDO DE BRUYNE_F
Nanyang Technol Univ, Ctr Signal Proc, Singapore 2263,;Singapore;
Univ Melbourne, Dept Elect & Elect Engn, Parkville, Vic 3052, Australia;
Australian Natl Univ, Dept Syst Engn, RSISE, Canberra, ACT 2601, Australia;
Siemens Brussels, Adv Proc Control Grp, Huizingen, Belgium

2: OTTERPOHL_JR EMMERT-STREIB_F PAWELZIK_K
Univ Bremen, Inst Theoret Neurophys, Ctr Cognit Sci, Kufsteiner;Str, D-28334 Bremen, Germany;
Univ Bremen, Inst Theoret Neurophys, Ctr Cognit Sci, D-28334 Bremen, Germany;
Univ Bremen, Ctr Cognit Sci, D-28334 Bremen, Germany

3: KSCHISCHANG_FR SOROKINE_V PASUPATHY_S
Qualcomm Inc, San Diego, CA 92121 USA;
Univ Toronto, Dept Elect & Comp Engn, Toronto, ON M5S 3G4, Canada

4: ZHANG_YY ZHANG_D ZHU_XY
Tsing Hua Univ, Dept Comp Sci, Beijing, Peoples R China;
Tsing Hua Univ, Dept Comp Sci, Beijing, Peoples R China;
Hong Kong Polytech Univ, Dept Comp, Hong Kong, Hong Kong

5: COX_RV KAMM_CA RABINER_LR SCHROETER_J WILPON_JG
AT&T Bell Labs, Speech & Image Proc Serv Res Lab, Florham Park, NJ 07932 USA

6: LEE_CH JUANG_BH Li_Q ZHOU_QR
AT&T BELL LABS,SPEECH RES DEPT,LUCENT TECHNOL,MURRAY HILL,NJ;07974

7: GAUVAIN_JL LAMEL_L ADDA_G
CNRS, LIMSI, Spoken Language Proc Grp, BP 133, F-91403 Orsay,;France;CNRS, LIMSI, Spoken Language Proc Grp, F-91403 Orsay, France

8: DEMBO_A ROSEN_J PERES_Y ZEITOUNI_O
Technion Israel Inst Technol, Dept Elect Engn, IL-32000 Haifa,;Israel;
Technion Israel Inst Technol, Dept Elect Engn, IL-32000 Haifa, Israel;
Stanford Univ, Dept Math, Stanford, CA 94305 USA;
Stanford Univ, Dept Math, Stanford, CA 94305 USA;
Univ Calif Berkeley, Dept Stat, Berkeley, CA 94720 USA;
Hebrew Univ Jerusalem, Inst Math, IL-91904 Jerusalem, Israel;
CUNY Coll Staten Isl, Dept Math, Staten Isl, NY 10314 USA

9: POORE_CA COKER_C MOBLEY_HLT
Univ Maryland, Sch Med, Dept Microbiol & Immunol, 655 W;Baltimore St,BRB 13-009, Baltimore, MD 21201 USA

10: COKER_C AYDINOK_Y KAVAKLI_K NISLI_G KANTAR_M CETINGUL_N
Ege Univ, Fac Med, Dept Pediat Hematol, Izmir, Turkey;
Ege Univ, Fac Med, Dept Pediat Hematol, Izmir, Turkey;
Ege Univ, Fac Med, Res Lab, Izmir, Turkey;
Ege Univ, Fac Med, Dept Pediat Oncol, Izmir, Turkey

11: MYERS_CS CONTRERAS_MA CHANg_MCJ RAPOPORT_SI GREINER_RS SALEM_N
Univ Maryland, Maryland Psychiat Res Ctr, POB 21247, Baltimore,;MD 21228 USA;
NIA, Sect Brain Physiol & Metab, NIH, Bethesda, MD USA;
US FDA, Ctr Drug Evaluat & Res, Off Testing & Res, Div Appl Pharmacol Res, Laurel, MD USA

12: SCHMIDT_CE HUDSON_TW EVANS_GRD
Univ Texas, Dept Chem Engn, Austin, TX 78712 USA;
Univ Texas, Dept Chem Engn, Austin, TX 78712 USA;
Univ Texas, Dept Biomed Engn, Austin, TX 78712 USA;
Univ Texas, MD Anderson Canc Ctr, Dept Plast Surg, Houston, TX 77030 USA

13: RUNKLE_P CARIN_L YODER_TJ BUCARO_JA
Duke Univ, Dept Elect & Comp Engn, Box 90291, Durham, NC 27708;USA;
Duke Univ, Dept Elect & Comp Engn, Durham, NC 27708 USA;
USN, Res Lab, Washington, DC 20375 USA;SFA Inc, Largo, MD 20785 USA

14: RUNKLE_P COUCHMAN_L CARIN_L
Duke Univ, Dept Elect & Comp Engn, Box 90291, Durham, NC 27708;USA;
Duke Univ, Dept Elect & Comp Engn, Durham, NC 27708 USA;
USN, Res Lab, Washington, DC 20375 USA;SFA Inc, Largo, MD 20785 USA

15: CARIN_L RUNKLE_PR BHARADWAJ_PK
Duke Univ, Dept Elect & Comp Engn, Durham, NC 27708 USA;
Duke Univ, Dept Elect & Comp Engn, Durham, NC 27708 USA

16: TUR_G HAKKANI-TUR_D STOLCKE_A SHRIBERG_E
SRI Int, Speech Technol & Res Lab, 333 Ravenswood Ave, Menlo;Park, CA 94025 USA;
SRI Int, Speech Technol & Res Lab, Menlo Park, CA 94025 USA;
Bilkent Univ, Dept Comp Engn, TR-06533 Ankara, Turkey

17: SCHWARTZ_R BAZZI_I MAKHOUL_J
Verizon, BBN Technol, Cambridge, MA 02138 USA;

18: MULLER_S EICKELER_S RIGOLL_G
Univ Duisburg Gesamthsch, Fac Elect Engn, Dept Comp Sci, D-4100;Duisburg, Germany;
Univ Duisburg Gesamthsch, Fac Elect Engn, Dept Comp Sci, D-4100 Duisburg, Germany

19: HAZOUT_S CAMPROUX_AC TUFFERY_P BOISVIEUX_JF
Univ Paris 07, INSERM, U155, Equipe Bioinformat Mol, Case;7113,2 Pl Jussieu, F-75251 Paris 05, France;
Univ Paris 07, INSERM, U155, Equipe Bioinformat Mol, F-75251 Paris 05, France;
CHU Pitie Salpetriere, Dept Biomath, F-75013 Paris, France

20: QIN_F AUERBACH_A SACHS_F
SUNY BUFFALO,DEPT BIOPHYS SCI,BUFFALO,NY 14214

21: THRUN_S BURGARD_W CREMERS_AB FOX_D HAHNEL_D SCHULZ_D
Aachen Tech Univ, Comp Sci Dept 5, Aachen, Germany
Carnegie Mellon Univ, Dept Comp Sci, Pittsburgh, PA 15213 USA;
Carnegie Mellon Univ, Dept Comp Sci, Pittsburgh, PA 15213 USA;
Carnegie Mellon Univ, Inst Robot, Pittsburgh, PA 15213 USA;
Univ Bonn, Comp Sci Dept 3, D-5300 Bonn, Germany

22: BOSSERT_M SHAVGULIDZE_S HAUTLE_A
Univ Ulm, Dept Informat Technol, D-89081 Ulm, Germany;
Univ Ulm, Dept Informat Technol, D-89081 Ulm, Germany;
Georgian Techn Univ, Dept Digital Commun Theory, GE-380075 Tbilisi, Rep of Georgia

23: DOGAN_S SADKA_AH KONDOZ_AM
Univ Surrey, Ctr Commun Syst Res, Guildford GU2 7XH, Surrey,;England;
Univ Surrey, Ctr Commun Syst Res, Guildford GU2 7XH, Surrey, England

24: MIN_BW YOON_HS SOH_J OHASHI_T EJIMA_T
ETRI, CSTL, Image Proc Dept, Yuseong Gu, 161 Kajong Dong,;Taejon 305350, South Korea;
ETRI, CSTL, Image Proc Dept, Yuseong Gu, Taejon 305350, South Korea;
Kyushu Inst Technol, Dept Artificial Intelligence, Iizuka, Fukuoka 820, Japan

ANNEX VII
Collaboration Groups identified by Co-authorship analysis in Language Research (1981 - 2001)

Publications during 1981-1990
1 Clique Found

1: RENIER_W ELING_P SLIS_I DEBOT_K
UNIV NIJMEGEN,DEPT EXPTL PSYCHOL,6500 HE NIJMEGEN,NETHERLANDS

Publications during 1981-1991
3 cliques found.

1: CROOKES_D MORROW_PJ MCPARLAND_PJ
QUEENS UNIV BELFAST,DEPT COMP SCI,BELFAST BT7 1NN,ANTRIM,NORTH;IRELAND

2: RENIER_W ELING_P SLIS_I DEBOT_K
UNIV NIJMEGEN,DEPT EXPTL PSYCHOL,6500 HE NIJMEGEN,NETHERLANDS

3: NUMAZAKI_H TAMURA_N TANAKA_H
YOKOHAMA NATL UNIV,YOKOHAMA,JAPAN;TOKYO INST TECHNOL,MEGURO KU,TOKYO 152,JAPAN

Publications during 1981 - 1992
14 cliques found.

1: WISE_R FRISTON_K FRACKOWIAK_R
HAMMERSMITH HOSP,MRC,CYCLOTRON UNIT,LONDON W12 0HS,ENGLAND
UNIV CAMBRIDGE,MRC,APPL PSYCHOL UNIT,CAMBRIDGE CB2 1TN,ENGLAND

2: WISE_R CHOLLET_F FRACKOWIAK_R
HAMMERSMITH HOSP,MRC,CYCLOTRON UNIT,LONDON W12 0HS,ENGLAND
UNIV CAMBRIDGE,MRC,APPL PSYCHOL UNIT,CAMBRIDGE CB2 1TN,ENGLAND

3: WISE_R FRACKOWIAK_R RAMSAY_S
HAMMERSMITH HOSP,MRC,CYCLOTRON UNIT,LONDON W12 0HS,ENGLAND
UNIV CAMBRIDGE,MRC,APPL PSYCHOL UNIT,CAMBRIDGE CB2 1TN,ENGLAND
NUCL MED & ULTRASOUND ASSOCIATES,ASHLEY CTR,WESTMEAD,NSW,AUSTRALIA

4: WISE_R HOWARD_D PATTERSON_K
HAMMERSMITH HOSP,MRC,CYCLOTRON UNIT,LONDON W12 0HS,ENGLAND
Inst Neurol, Wellcome Dept Cognit Neurol, London WC1N 3BG, England
UNIV LONDON BIRKBECK COLL,DEPT PSYCHOL,LONDON WC1E 7HX,ENGLAND

5: CROOKES_D MORROW_PJ MCPARLAND_PJ
QUEENS UNIV BELFAST,DEPT COMP SCI,BELFAST BT7 1NN,ANTRIM,NORTH;IRELAND

6: RENIER_W ELING_P SLIS_I DEBOT_K
UNIV NIJMEGEN,DEPT EXPTL PSYCHOL,6500 HE NIJMEGEN,NETHERLANDS

7: IKEDA_T HERATH_S ISHIZAKI_S ANZAI_Y AISO_H
SW TEXAS STATE UNIV,SAN MARCOS,TX 78666

8: HEERING_J KLINT_P REKERS_J

CENT WISKUNDE INFORMAT,POB 4079,1009 AB AMSTERDAM,NETHERLANDS;
UNIV AMSTERDAM,AMSTERDAM,NETHERLANDS

9: SUMITA_K UKITA_T AMANO_S
TOSHIBA CO LTD,CTR RES & DEV,KAWASAKI 210,JAPAN

10: BROWN_HK HAZELTON_TR SILBIGER_ML
UNIV S FLORIDA,COLL MED,DEPT ANAT,TAMPA,FL 33612;
UNIV S FLORIDA,COLL MED,DEPT RADIOL,TAMPA,FL 33612

11: BECHTEREVA_NP ABDULLAEV_YG MEDVEDEV_SV
INST EXPTL MED,DEPT HUMAN NEUROPHYSIOL,LENINGRAD,USSR

12: NUMAZAKI_H TAMURA_N TANAKA_H
YOKOHAMA NATL UNIV,YOKOHAMA,JAPAN;TOKYO INST TECHNOL,MEGURO KU,TOKYO 152,JAPAN

13: QU_DL MURAKI_M HAYAKAWA_T
TOKYO INST TECHNOL,DEPT IND ENGN & MANAGEMENT,TOKYO 152,JAPAN

14: LANE_TA ANDERSON_KC GOODNOUGH_LT SILBERSTEIN_LE
UNIV CALIF SAN DIEGO,SCH MED,DEPT PATHOL 0612,9500 GILMAN DR,LA;JOLLA,CA 92093;
HARVARD UNIV,SCH MED,DANA FARBER CANC INST,BOSTON,MA 02115;
UNIV HOSP CLEVELAND,DEPT MED,CLEVELAND,OH 44106
HOSP UNIV PENN,BLOOD BANK,PHILADELPHIA,PA 19104

Publications during 1993 - 1994
8 cliques found.

1: RUMSEY_JM ZAMETKIN_AJ ANDREASON_P HANAHAN_AP HAMBURGER_SD AQUINO_T KING_AC PIKUS_A COHEN_RM
NIMH,CHILD PSYCHIAT BRANCH,BLDG 10,ROOM 6N240,BETHESDA,MD 20892;
NIMH,CEREBRAL METAB LAB,CLIN BRAIN IMAGING SECT,BETHESDA,MD 20892;
NIDCD,CLIN AUDIOL HEARING SECT,BETHESDA,MD

2: FRIEDMAN_C CIMINO_JJ JOHNSON_SB
CUNY,QUEENS COLL,NEW YORK,NY;COLUMBIA UNIV,NEW YORK,NY

3: TOMABECHI_H MAEDA_AM AOE_JI
UNIV TOKUSHIMA,DEPT INFORMAT SCI & INTELLIGENT SYST,2-1 MINAMI;JOSANJIMA CHO,TOKUSHIMA 770,JAPAN

4: SAGER_N LYMAN_M TICK_LJ
NYU,COURANT INST MATH SCI,251 MERCER ST,NEW YORK,NY 10012;NYU,SCH MED,NEW YORK,NY 10012

5: DANIELE_A GIUSTOLISI_L SILVERI_MC
CATHOLIC UNIV SACRED HEART,IST NEUROL,LARGO A GEMELLI 8,I-00168;ROME,ITALY

6: FITCH_RH TALLAL_P BROWN_CP
RUTGERS STATE UNIV,CTR MOLEC & BEHAV NEUROSCI,197 UNIV;AVE,NEWARK,NJ 07102

7: MCCARTHY_G BLAMIRE_AM ROTHMAN_DL SHULMAN_RG
YALE UNIV,SCH MED,DEPT SURG NEUROSURG,NEW HAVEN,CT 06510;

YALE UNIV,SCH MED,DEPT MOLEC BIOPHYS & BIOCHEM,NEW HAVEN,CT 06510;
YALE UNIV,SCH MED,DEPT INTERNAL MED,NEW HAVEN,CT 06510;
VET ADM MED CTR,NEUROPSYCHOL LAB 116B1,W HAVEN,CT 06516

8: TANAKA_H SHIMADA_K KOIKE_H
UNIV TOKYO,FAC ENGN,DEPT ELECT ENGN,7-3-1 HONGO,BUNKYO KU,TOKYO;113,JAPAN

Publications in 1995 - 1996
35 cliques found.

1: PUGH_KR SHAYWITZ_BA SHAYWITZ_SE CONSTABLE_RT SKUDLARSKI_P FULBRIGHT_RK BRONEN_RA SHANKWEILER_DP KATZ_L FLETCHER_JM GORE_JC
YALE UNIV,SCH MED,DEPT PEDIAT,POB 3333,NEW HAVEN,CT 06510;
YALE UNIV,SCH MED,DEPT NEUROL,NEW HAVEN,CT 06510;
YALE UNIV,SCH MED,DEPT DIAGNOST RADIOL,NEW HAVEN,CT 06510;
YALE UNIV,HASKINS LABS INC,NEW HAVEN,CT 06510;
YALE UNIV,DEPT APPL PHYS,NEW HAVEN,CT 06510;
UNIV TEXAS,SCH MED,DEPT PEDIAT,HOUSTON,TX;
UNIV CONNECTICUT,DEPT PSYCHOL,STORRS,CT 06269

2: SHAYWITZ_BA SHAYWITZ_SE SHANKWEILER_DP KATZ_L FLETCHER_JM MARCHIONE_KE
YALE UNIV,SCH MED,DEPT PEDIAT,NEW HAVEN,CT 06520;
YALE UNIV,SCH MED,DEPT NEUROL,NEW HAVEN,CT 06520;
YALE UNIV,SCH MED,HASKINS LABS,NEW HAVEN,CT 06520;
YALE UNIV,SCH MED,DEPT DIAGNOST RADIOL,NEW HAVEN,CT 06520;
UNIV TEXAS,SCH MED,DEPT PEDIAT,HOUSTON,TX;
YALE UNIV,DEPT APPL PHYS,NEW HAVEN,CT 06520

3: PERANI_D DEHAENE_S GRASSI_F COHEN_L CAPPA_SF DUPOUX_E FAZIO_F MEHLER_J
UNIV MILAN,CNR,INB,DIPSCO,SCI INST HS RAFFAELE,VIA OLGETTINA;60,I-20132 MILAN,ITALY;
UNIV BRESCIA,INB,CNR,DIPSCO,SCI INST HS RAFFAELE,I-20132 MILAN,ITALY;
CNRS,EHESS,LSCP,PARIS,FRANCE

4: EDEN_GF VANMETER_JW RUMSEY_JM ZEFFIRO_TA
NIMH,SECT FUNCT BRAIN IMAGING,NIH,BETHESDA,MD 20892;
NIMH,CHILD PSYCHIAT BRANCH,NIH,BETHESDA,MD 20892;
NIH,LAB DIAGNOST RADIOL RES,OD,BETHESDA,MD 20892;
SENSOR SYST INC,STERLING,VA 20164;
UNIV CALIF LOS ANGELES,SCH MED,DEPT NEUROL,LOS ANGELES,CA 90098

5: BASILI_R PAZIENZA_MT VELARDI_P
UNIV ROMA TOR VERGATA,DIPARTIMENTO INFORMAT SISTEMI & PROD,VIA;RIC SCI,I-00133 ROME,ITALY;
ENEA,CTR STUDI & DOCUMENTAZ,CASACCIA,ITALY

6: BERNDT_RS HAENDIGES_AN MITCHUM_CC
UNIV MARYLAND,SCH MED,DEPT NEUROL,22 S GREENE ST,BALTIMORE,MD;21201

7: PULVERMULLER_F PREISSL_H LUTZENBERGER_W BIRBAUMER_N
UNIV TUBINGEN,INST MED PSYCHOL & VERHALTENSNEUROBIOL,GARTENSTR;29,D-72074 TUBINGEN,GERMANY

8: PULVERMULLER_F MOHR_B RAYMAN_J

UNIV TUBINGEN,INST MED PSYCHOL &
VERHALTENSNEUROBIOL,GARTENSTR;29,D-72074 TUBINGEN,GERMANY;
UNIV MUNSTER,INST EXPTL AUDIOL,D-48149 MUNSTER,GERMANY;
UNIV PADUA,PADUA,ITALY

9: JAGIRDAR_R JAIN_VK BATRA_JL DHANDE_SG
INDIAN INST TECHNOL,DEPT MECH ENGN,KANPUR
208016,UTTAR;PRADESH,INDIA

10: PRICE_CJ WISE_RJS FRACKOWIAK_RSJ
HAMMERSMITH HOSP,MRC,CYCLOTRON UNIT,DU CANE RD,LONDON
W12;0HS,ENGLAND;
CHARING CROSS HOSP,CTR NEUROSCI,LONDON W6 8RP,ENGLAND;
UNIV LONDON BIRKBECK COLL,DEPT PSYCHOL,LONDON WC1E 7HX,ENGLAND;
MRC,APPL PSYCHOL UNIT,CAMBRIDGE,ENGLAND

11: BINDER_JR HAMMEKE_TA FROST_JA RAO_SM
MED COLL WISCONSIN,DEPT NEUROL,9200 W WISCONSIN;AVE,MILWAUKEE,WI
53226;MED COLL WISCONSIN,DEPT BIOPHYS,MILWAUKEE,WI 53226

12: SWANSON_SJ HAMMEKE_TA MORRIS_GL MUELLER_WM HAUGHTON_VM
MED COLL WISCONSIN,DEPT NEUROSURG,MILWAUKEE,WI 53226;MED COLL
WISCONSIN,DEPT RADIOL,MILWAUKEE,WI 53226

13: DESPOSITO_M GROSSMAN_M ONISHI_K BIASSOU_N WHITEDEVINE_T
ROBINSON_KM
UNIV PENN,MED CTR,DEPT NEUROL,PHILADELPHIA,PA 19104

14: DESPOSITO_M GROSSMAN_M HUGHES_E ONISHI_K BIASSOU_N
ROBINSON_KM
UNIV PENN,MED CTR,DEPT NEUROL,PHILADELPHIA,PA 19104

15: GUNTER_TC JACKSON_JL MULDER_G
MAX PLANCK INST COGNIT NEUROSCI,INSELSTR 22-26,D-
04103;LEIPZIG,GERMANY;
UNIV GRONINGEN,DEPT LINGUIST,NL-9700 AB GRONINGEN,NETHERLANDS;
UNIV GRONINGEN,INST EXPT & OCCUPAT PSYCHOL,NL-9700 AB
GRONINGEN,NETHERLANDS

16: SUBRAHMANIAN_VS NERODE_A NG_RT
CORNELL UNIV,INST MATH SCI,ITHACA,NY 14853;
UNIV BRITISH COLUMBIA,DEPT COMP SCI,VANCOUVER,BC,CANADA;
UNIV MARYLAND,INST ADV COMP STUDIES,COLLEGE PK,MD 20742;
UNIV MARYLAND,DEPT COMP SCI,COLLEGE PK,MD 20742

17: SPITZER_M BELLEMANN_ME KAMMER_T GUCKEL_F SCHWARTZ_A BRIX_G
UNIV HEIDELBERG,PSYCHIAT KLIN,SEKT EXPT PSYCHOPATHOL,D-
6900;HEIDELBERG,GERMANY;
UNIV HEIDELBERG,KLINIKUM MANNHEIM,KLIN FAK 2,INST KLIN
RADIOL,HEIDELBERG,GERMANY;
UNIV HEIDELBERG,KLINIKUM MANNHEIM,KLIN FAK 2,NEUROL
KLIN,HEIDELBERG,GERMANY;
UNIV MAGDEBURG,NEUROL KLIN,D-39106 MAGDEBURG,GERMANY;
DEUTSCH KREBSFORSCHUNGSZENTRUM,FORSCH SCHWERPUNKT RADIOL
DIAGNOT & THERAPIE,D-6900 HEIDELBERG,GERMANY

18: BATES_E DEVESCOVI_A HERNANDEZ_A PIZZAMIGLIO_L
UNIV CALIF SAN DIEGO,CTR RES LANGUAGE 0526,LA JOLLA,CA 92093;

UNIV ROMA LA SAPIENZA,ROME,ITALY;UNIV CALIF SAN DIEGO,LA JOLLA,CA 92093;
UNIV ROMA LA SAPIENZA,CLIN SANTA LUCIA,ROME,ITALY

19: EULITZ_C PANTEV_C FEIGE_B ELBERT_T
Univ Konstanz, Fachgrp Psychol, Postfach 5560-D25, D-78434;Constance, Germany;
Univ Konstanz, Fachgrp Psychol, D-78434 Constance, Germany;
Zentrum Psychiat, Reichenau, Germany

20: KARBE_H HALBER_M HEISS_WD
UNIV COLOGNE,MAX PLANCK INST NEUROL FORSCH,W-5000;COLOGNE,GERMANY;
UNIV COLOGNE,NEUROL KLIN,W-5000 COLOGNE,GERMANY

21: TRAUM_DR SCHUBERT_LK HWANG_CH HEEMAN_P FERGUSON_G ALLEN_JF POESIO_M LIGHT_M
UNIV GENEVA,FPSE,TECFA,9 RTE DR,CH-1227 CAROUGE,SWITZERLAND;
UNIV ROCHESTER,ROCHESTER,NY 14627;
UNIV EDINBURGH,EDINBURGH EH8 9YL,MIDLOTHIAN,SCOTLAND;
UNIV TUBINGEN,TUBINGEN,GERMANY

22: TALLAL_P MILLER_SL SCHREINER_C JENKINS_WM MERZENICH_MM
UNIV CALIF SAN FRANCISCO,WM KECK CTR INTEGRAT NEUROSCI,SAN;FRANCISCO,CA 94143;
UNIV CALIF SAN FRANCISCO,COLEMAN LAB,SAN FRANCISCO,CA 94143;
RUTGERS STATE UNIV,CTR MOLEC & BEHAV NEUROSCI,NEWARK,NJ 07102

23: GROSSMAN_M HUGHES_E ONISHI_K MICKANIN_J
UNIV PENN,MED CTR,DEPT NEUROL,PHILADELPHIA,PA 19104

24: CHIANG_TH LIN_YC SU_KY
IND TECHNOL RES INST,CTR ADV TECHNOL,COMP & COMMUN RES;LABS,HSINCHU 310,TAIWAN

25: EDGINGTON_M LOWRY_A JACKSON_P BREEN_AP MINNIS_S
BT LABS,SPEECH SYNTH & ANAL GRP,IPSWICH,SUFFOLK,ENGLAND

26: CORINA_D BAVELIER_D JEZZARD_P CLARK_V PADMANHABAN_S RAUSCHECKER_J TURNER_R NEVILLE_H
UNIV WASHINGTON,SEATTLE,WA 98195;NIMH,NIH,ROCKVILLE,MD;
UNIV OREGON,EUGENE,OR 97403;
UNIV MILAN,INB,CNR,INST HS RAFFAELE,MILAN,ITALY;
NIDCD,NIH,BETHESDA,MD;
EPSTEIN LABS,SAN FRANCISCO,CA

27: FERRUCCI_F TORTORA_G TUCCI_M VITIELLO_G
Univ Salerno, Dipartimento Informat & Applicaz, I-84081;Baronissi, Salerno, Italy

28: HICKOK_G BELLUGI_U KLIMA_ES
SALK INST BIOL STUDIES,COGNIT NEUROSCI LAB,10010 N TORREY PINES;RD,LA JOLLA,CA 92037;
UNIV CALIF SAN DIEGO,SAN DIEGO,CA 92103;
SAN DIEGO VET ADM HOSP,LA JOLLA,CA 92093

29: JUST_MA CARPENTER_PA MIYAKE_A
CARNEGIE MELLON UNIV,DEPT PSYCHOL,PITTSBURGH,PA 15213

30: GIUNCHIGLIA_E ARMANDO_A TRAVERSO_P CIMATTI_A

UNIV GENOA,DIST,VIA OPERA PIA 11A,I-16145 GENOA,ITALY;
IRST,LOCALITA PANTE POVO,I-38050 TRENT,ITALY;
UNIV ANCONA,ANCONA,ITALY

31: SHIFFMAN_S DETMER_WM LANE_CD FAGAN_LM
STANFORD UNIV,MED INFORMAT SECT,MED SCH OFF BLDG;X215,STANFORD,CA 94305

32: RASSINOUX_AM JUGE_C MICHEL_PA BAUD_RH DEGOULET_P SCHERRER_JR
UNIV GENEVA,FAC MED,GENEVA,SWITZERLAND;BROUSSAIS UNIV HOSP,PARIS,FRANCE

33: MICHEL_PA SCHERRER_JR LOVIS_C
UNIV STATE HOSP GENEVA,DEPT INTERNAL MED,GENEVA,SWITZERLAND;
UNIV STATE HOSP GENEVA,DIV MED INFORMAT,GENEVA,SWITZERLAND;
VANDERBILT UNIV,DIV BIOMED INFORMAT,NASHVILLE,TN

34: BAUD_RH SCHERRER_JR WAGNER_JC
UNIV HOSP GENEVA,DEPT MED INFORMAT,GENEVA,SWITZERLAND

35: MUNTE_TF MATZKE_M JOHANNES_S
HANNOVER MED SCH,NEUROL KLIN KLIN NEUROPHYSIOL,HANNOVER,GERMANY

Publications during 1997
7 cliques found.

1: SHULMAN_GL CORBETTA_M BUCKNER_RL FIEZ_JA MIEZIN_FM RAICHLE_ME PETERSEN_SE
UNIV WASHINGTON,SCH MED,SEATTLE,WA 98195

2: ENGELS_G HECKEL_R TAENTZER_G EHRIG_H
Leiden Univ, Dept Comp Sci, POB 9512, NL-2300 RA Leiden,;Netherlands;
Leiden Univ, Dept Comp Sci, NL-2300 RA Leiden, Netherlands;
Tech Univ Berlin, Dept Comp Sci, D-10587 Berlin, Germany

3: KACSUK_P DOZSA_G FADGYAS_T
HUNGARIAN ACAD SCI,RES INST MEASUREMENT & COMP;TECHNIQUES,KFKI,MSZKI,POB 49,H-1525 BUDAPEST,HUNGARY

4: DELISI_LE SAKUMA_M KUSHNER_M FINER_DL HOFF_AL
SUNY STONY BROOK,DEPT PSYCHIAT,HSC,T-10,STONY BROOK,NY 11794;
SUNY STONY BROOK,SCHIZOPHRENIA RES PROJECT,STONY BROOK,NY 11794;
SUNY STONY BROOK,DEPT LINGUIST,STONY BROOK,NY 11794;NAPA STATE HOSP,NAPA,CA;
UNIV CALIF DAVIS,DAVIS,CA

5: TEICH_E HAGEN_E GROTE_B
GERMAN NATL CTR INFORMAT TECHNOL GMD,D-64293 DARMSTADT,GERMANY;
UNIV SAARLAND,D-6600 SAARBRUCKEN,GERMANY;
TECH UNIV DARMSTADT,D-64287 DARMSTADT,GERMANY

6: SHELTON_JR WEINRICH_M MCCALL_D COX_DM
UNIV MARYLAND,DEPT NEUROL & REHABIL,BALTIMORE,MD 21201

7: NAATANEN_R CHEOUR_M ALHO_K

UNIV HELSINKI,DEPT PSYCHOL,COGNIT BRAIN RES UNIT,FIN-00014;HELSINKI,FINLAND
UNIV HELSINKI,DEPT PHONET,FIN-00014 HELSINKI,FINLAND
TARTU STATE UNIV,DEPT PSYCHOL,EE-202400 TARTU,ESTONIA

Publications during 1998
12 cliques found.

1: DE SPYNS_P MOOR_G CEUSTERS_W
Zonnegem, Belgium;Language & Comp NV, B-9520 Zonnegem, Belgium;
State Univ Ghent Hosp, RAMIT VZW, B-9000 Ghent, Belgium
Language & Comp NV, Het Moorhof,Hazenakkerstr 20, B-9520;

3: BAUD_RH RASSINOUX_AM SCHERRER_JR
UNIV STATE HOSP GENEVA,CTR INFORMAT HOSP,CH-1211 GENEVA;4,SWITZERLAND

4: LOVIS_C RASSINOUX_AM SCHERRER_JR
UNIV STATE HOSP GENEVA,DEPT MED,GENEVA,SWITZERLAND;
UNIV STATE HOSP GENEVA,INFORMAT CTR,GENEVA,SWITZERLAND

5: VON FRIEDERICI_AD HAHNE, A CRAMON_DY
Max Planck Inst Cognit Neurosci, Inselstr 22-26, D-04103;Leipzig, Germany;
Max Planck Inst Cognit Neurosci, D-04103 Leipzig, Germany

6: KARBE_H HERHOLZ_K HEISS_WD
Max Planck Inst Neurol Forsch, Gleueler Str 50, D-50931;Cologne, Germany
Univ Hosp, Dept Neurol, Cologne, Germany

7: NEVILLE_HJ BAVELIER_D CORINA_D KARNI_A LALWANI_A BRAUN_A CLARK_V JEZZARD_P TURNER_R
Univ Oregon, Dept Psychol, Eugene, OR 97403 USA;
Univ Oregon, Dept Psychol, Eugene, OR 97403 USA;
Georgetown Univ, Med Ctr, Washington, DC 20007 USA;
Univ Washington, Seattle, WA 98195 USA;Weizmann Inst Sci, IL-76100 Rehovot, Israel;NIH, Bethesda, MD 20892 USA;
Inst Neurol, London WC1N 3BG, England;
Univ Calif San Francisco, San Francisco, CA 94143 USA

8: GOLDBERG_TE ALOIA_MS GOUROVITCH_ML MISSAR_D PICKAR_D WEINBERGER_DR
NIMH, Clin Brain Disorders Branch, Bldg 10,Rm 4S235,MSC 1379,;Bethesda, MD 20892 USA;
NIMH, Clin Brain Disorders Branch, Bethesda, MD 20892 USA;
NIMH, Expt Therapeut Branch, Bethesda, MD 20892 USA

9: SPITZER_M KAMMER_T SEYYEDI_S
Univ Ulm, Abt Psychiat 3, Psychiat Klin, D-89075 Ulm, Germany
UNIV HEIDELBERG,KLINIKUM MANNHEIM,KLIN FAK 2,NEUROL KLIN,HEIDELBERG,GERMANY
Max Planck Inst Biol Kybernet, D-72072 Tubingen, Germany

10: MACKAY_DG STEWART_R BURKE_DM
Univ Calif Los Angeles, Dept Psychol, Los Angeles, CA 90095 USA;
Univ Calif Los Angeles, Dept Psychol, Los Angeles, CA 90095 USA;
Pomona Coll, Claremont, CA 91711 USA

11: VARGHA-KHADEM_F WATKINS_KE PEMBREY_ME
Univ Coll London, Sch Med, Inst Child Hlth, Cognit Neurosci;Unit,Wolfson Ctr, Mecklenburgh Sq, London WC1N 2AP, England

12: OVIATT_S LEVOW_GA MACEACHERN_M
Oregon Grad Inst Sci & Technol, Dept Comp Sci, Ctr Human Comp;Commun, POB 91000, Portland, OR 97291 USA
Oregon Grad Inst Sci & Technol, Dept Comp Sci, Ctr Human Comp Commun, Portland, OR 97291 USA
Univ Pittsburgh, Dept Linguist, Pittsburgh, PA 15260 USA;MIT, Artificial Intelligence Lab, Cambridge, MA 02139 USA

13: KNECHT_S DEPPE_M HENNINGSEN_H HUBER_T RINGELSTEIN_EB
Univ Munster, Dept Neurol, Albert Schweitzer Str 33, D-48129;Munster, Germany

Publications in 1999
23 cliques found.

1: VAN BERKUM_JJA BROWN_CM HAGOORT_P
Max Planck Inst Psycholinguist, POB 310, NL-6500 AH Nijmegen,;Netherlands;
Max Planck Inst Psycholinguist, NL-6500 AH Nijmegen, Netherlands

2: DAELEMANS_W VAN DEN BOSCH_A WEIJTERS_T
Tilburg Univ, ILK Res Grp, NL-5000 LE Tilburg, Netherlands
Univ Maastricht, MATRIKS, Dept Comp Sci, Maastricht, Netherlands

3: COHEN_L VAN DE MOORTELE_PF DEHAENE_S
CEA, DSV, INSERM U334, Serv Hosp Frederic Joliot, F-91401 Orsay, France;
Hop La Pitie Salpetriere, Paris, France

4: FUNNELL_MG CORBALLIS_PM GAZZANIGA_MS
Dartmouth Coll, Ctr Cognit Neurosci, Hanover, NH 03755 USA;

5: GLOVER_GH DESMOND_JE GABRIELI_JDE WAGNER_AD
Stanford Univ, Dept Radiol, Stanford, CA 94305 USA

6: STAAB_S BRAUN_C BRUDER_I DUSTERHOFT_A HEUER_A KLETTKE_M NEUMANN_G PRAGER_B PRETZEL_J SCHNURR_HP STUDER_R WRENGER_B
Univ Karlsruhe, AIFB, Kaiserstr 12, D-76128 Karlsruhe, Germany;Univ Karlsruhe, AIFB, D-76128 Karlsruhe, Germany;
DFKI, D-66123 Saarbrucken, Germany;
GECKO MBH, D-18055 Rostock, Germany;
Univ Rostock, Fachbereich Informat, D-18051 Rostock, Germany

7: LEVELT_WJM ROELOFS_A MEYER_AS
Max Planck Inst Psycholinguist, Wundtlaan 1, NL-6500 AH;Nijmegen, Netherlands;Max Planck Inst Psycholinguist, NL-6500 AH Nijmegen, Netherlands;
Univ Exeter, Sch Psychol, Washington Singer Labs, Exeter EX4 4QG, Devon, England

8: CROSSON_B SADEK_JR GOKCAY_D MARON_L AUERBACH_EJ BROWD_SR BRIGGS_RW
Univ Florida, Hlth Sci Ctr, Dept Clin & Hlth Psychol, Box;100165, Gainesville, FL 32610 USA;
Univ Florida, Hlth Sci Ctr, Dept Clin & Hlth Psychol, Gainesville, FL 32610 USA;
Univ Florida, Hlth Sci Ctr, Dept Comp Sci & Informat Engn, Gainesville, FL 32610 USA;
Univ Florida, Hlth Sci Ctr, Dept Nucl & Radiol Engn, Gainesville, FL 32610 USA;

Univ Florida, Hlth Sci Ctr, Dept Neurosci, Gainesville, FL 32610 USA;Univ Florida, Hlth Sci Ctr, Dept Radiol, Gainesville, FL 32610 USA; Univ Florida, Inst Brain, Gainesville, FL 32610 USA

9: MOHR_CM LEONARD_CM FREEMAN_AJ BRIGGS_RW
Univ Florida, Dept Nucl & Radiol Engn Sci, POB 100245,;Gainesville, FL 32610 USA;
Univ Florida, Dept Nucl & Radiol Engn Sci, Gainesville, FL 32610 USA;
Univ Florida, Dept Commun Sci & Disorders, Gainesville, FL 32611 USA;SMIS Ltd, Guildford GU23 5YF, Surrey, England;Univ Florida, Dept Radiol, Gainesville, FL 32610 USA;Univ Florida, Dept Chem, Gainesville, FL 32610 USA;Univ Florida, Dept Mol Biol & Biochem, Gainesville, FL 32610 USA;
Univ Florida, Dept Neurosci, Gainesville, FL 32610 USA

10: BULLMORE_ET BRAMMER_MJ WILLIAMS_SCR MURRAY_RM MCGUIRE_PK
INST PSYCHIAT,LONDON SE5 8AF,ENGLAND

11: CHEE_MWL TAN_EWL THIEL_T
Singapore Gamma Knife Ctr, 20 Coll Rd, Singapore 169856,;Singapore;
Singapore Gen Hosp, Cognit Neurosci Lab, Singapore 169856, Singapore;
Univ Freiburg, Dept Radiol, D-79104 Freiburg, Germany

12: SKRANDIES_W REIK_P KUNZE_C
Univ Giessen, Inst Physiol, Aulweg 129, D-35392 Giessen,;Germany

13: MULLER_RA BEHEN_ME ROTHERMEL_RD MUZIK_O CHAKRABORTY_PK CHUGANI_HT
Childrens Hosp, Res Ctr, 8110 La Jolla Shores Dr 200, La Jolla,;CA 92037 USA;
Wayne State Univ, Med Ctr, Dept Pediat, Detroit, MI 48202 USA;
Wayne State Univ, Med Ctr, Dept Psychiat, Detroit, MI 48202 USA;
Wayne State Univ, Med Ctr, Dept Radiol, Detroit, MI 48202 USA;
Wayne State Univ, Med Ctr, Dept Neurol, Detroit, MI 48202 USA;
Univ Calif San Diego, Dept Cognit Sci, San Diego, CA 92103 USA

14: MCCARLEY_RW HIRAYASU_Y SHENTON_ME
Harvard Univ, Harvard Brockton VAMC, Sch Med, Dept Psychiat,;940 Belmont St, Brockton, MA 02401 USA

15: MCCARLEY_RW NIZNIKIEWICZ_MA SHENTON_ME
Harvard Univ, Harvard Brockton VAMC, Sch Med, Dept Psychiat,;940 Belmont St, Brockton, MA 02401 USA

16: SERAFINI_S STEURY_K CORINA_D POSSE_S
Univ Washington, Dept Speech & Hearing Sci, Seattle, WA 98195;USA;
Univ Washington, Dept Psychol, Seattle, WA 98195 USA;Univ Washington, Dept Radiol, Seattle, WA 98195 USA;Univ Washington, Dept Psychiat, Seattle, WA 98195 USA;
Univ Washington, Dept Educ, Seattle, WA 98195 USA

17: BINDER_JR FROST_JA HAMMEKE_TA BELLGOWAN_PSF RAO_SM COX_RW
Med Coll Wisconsin, Dept Neurol, 9200 W Wisconsin Ave,;Milwaukee, WI 53226 USA;
Med Coll Wisconsin, Dept Neurol, Milwaukee, WI 53226 USA;
Med Coll Wisconsin, Dept Cellular Biol & Anat, Milwaukee, WI 53226 USA;
Med Coll Wisconsin, Biophys Res Inst, Milwaukee, WI 53226 USA

18: HAHN_U ROMACKER_M SCHULZ_S
Univ Freiburg, Computat Linguist Lab, Werthmannpl 1, D-79085;
Freiburg, Germany;Univ Freiburg, Computat Linguist Lab, D-79085 Freiburg, Germany;
Freiburg Univ Hosp, Dept Med Informat, D-79104 Freiburg, Germany

19: PATTERSON_K HODGES_JR CROOT_K
MRC, Cognit & Brain Sci Unit, 15 Chaucer Rd, Cambridge CB2 2EF,;England;

MRC, Cognit & Brain Sci Unit, Cambridge CB2 2EF, England;
Univ Cambridge, Addenbrookes Hosp, Neurol Unit, Cambridge CB2 2QQ, England
20: BROWN_CM HAGOORT_P TER KEURS_M
Max Planck Inst Psycholinguist, Wundtlaan 1, NL-6525 XD;Nijmegen, Netherlands;
Max Planck Inst Psycholinguist, NL-6525 XD Nijmegen, Netherlands

21: WACTLAR_HD CHRISTEL_MG HAUPTMANN_AG GONG_YH
Carnegie Mellon Univ, Pittsburgh, PA 15213 USA;Carnegie Mellon Univ, Pittsburgh, PA 15213 USA

22: BREIER_JI SIMOS_PG ZOURIDAKIS_G PAPANICOLAOU_AC
Univ Texas, Sch Med, Dept Neurosurg, 6431 Fannin,Suite 7-148,;Houston, TX 77030 USA

23: NIEMANN_H NOTH_E GALLWITZ_F
Univ Erlangen Nurnberg, Chair Pattern Recognit, Martensstr 3,;D-91058 Erlangen, Germany;
Univ Erlangen Nurnberg, Chair Pattern Recognit, D-91058 Erlangen, Germany;
Bavarian Res Ctr Knowledge Based Syst FORWISS, D-91058 Erlangen, Germany

Publications in 2000-2001
19 cliques found.
1: LE DEHAENE_S COHEN_L LEHERICY_S
Hop La Pitie Salpetriere, Serv Neurol 1, Clin Paul Castaigne, F-75651 Paris 13, France
Hop La Pitie Salpetriere, Serv Neurol 1, Clin Paul Castaigne,;47-83 Bd Hop, F-75651 Paris 13, France
EHESS, CNRS, Lab Sci Cognit & Psycholinguist, F-75651 Paris 13, France

2: LE COHEN_L LEHERICY_S BIHAN_D
Hop La Pitie Salpetriere, Serv Neurol 1, Clin Paul Castaigne,;47-83 Bd Hop, F-75651 Paris 13, France
EHESS, CNRS, Lab Sci Cognit & Psycholinguist, F-75651 Paris 13, France
CEA, Serv Hosp Frederic Joliot, DSV, INSERM,U334, F-91401 Orsay, France

3: VAN DE LE

4: COHEN_H LE NORMAND_MT
UQAM, Ctr Neurosci Cognit, Montreal, PQ, Canada

5: TUR_G HAKKANI-TUR_D STOLCKE_A SHRIBERG_E
SRI Int, Speech Technol & Res Lab, 333 Ravenswood Ave, Menlo;Park, CA 94025 USA;
SRI Int, Speech Technol & Res Lab, Menlo Park, CA 94025 USA;Bilkent Univ, Dept Comp Engn, TR-06533 Ankara, Turkey

6: PAPATHANASSIOU_D ETARD_O MELLET_E MAZOYER_B TZOURIO-MAZOYER_N
Univ Caen & CEA LRC 13V, UPRES EA 2127, GIN, GIP Cyceron, F-14074 Caen, France

7: DEPPE_M KNECHT_S LOHMANN_H HENNINGSEN_H DRAGER_B FLOEL_A
Univ Munster, Dept Neurol, Albert Schweitzer Str 33, D-48129;Munster, Germany

8: DEPPE_M KNECHT_S LOHMANN_H RINGELSTEIN_EB HENNINGSEN_H
Univ Munster, Dept Neurol, Albert Schweitzer Str 33, D-48129;Munster, Germany

9: FRIEDERICI_AD MEYER_M VON CRAMON_DY
Max Planck Inst Cognit Neurosci, D-04103 Leipzig, Germany

10: GAO_JH LIU_HL PU_YL FOX_PT
Univ Texas, Hlth Sci Ctr, Res Imaging Ctr, 7703 Floyd Curl Dr,;San Antonio, TX 78284 USA;
Univ Texas, Hlth Sci Ctr, Res Imaging Ctr, San Antonio, TX 78284 USA;

Chang Gung Univ, Dept Med Technol, Tao Yuan, Taiwan;
Chang Gung Med Ctr, Dept Diagnost Radiol, Tao Yuan, Taiwan;
Shandong Med Univ Hosp, Dept Radiol, Jinan, Peoples R China

11: XIONG_JH FOX_PT ZAMARRIPA_F
Univ Texas, Hlth Sci Ctr, Res Imaging Ctr, 7703 Floyd Curl;Dr,MSC 6240, San Antonio, TX 78229 USA
Univ Calif Santa Barbara, Dept Speech & Hearing Sci, Santa Barbara, CA 93106 USA

12: NAKAMURA_K HONDA_M OKADA_T HANAKAWA_T FUKUYAMA_H SHIBASAKI_H
Kyoto Univ, Grad Sch Med, Dept Brain Pathophysiol, Sakyo Ku, 54;
Shogoin, Kyoto 6068507, Japan;
Kyoto Univ, Grad Sch Med, Dept Brain Pathophysiol, Sakyo Ku, Kyoto 6068507, Japan;
Kyoto Univ, Grad Sch Med, Dept Nucl Med, Kyoto 6068507, Japan;
Natl Inst Physiol Sci, Lab Cerebral Integrat, Kyoto, Japan

13: GOLDBERG_TE EGAN_MF WEINBERGER_DR
NIMH, IRP, Clin Brain Disorders Branch, 10 Ctr Dr,MSC 1379,;Bethesda, MD 20892 USA;NIMH, IRP, Clin Brain Disorders Branch, Bethesda, MD 20892 USA

14: ARGYRIS_J DE DONNO_M LITVIN_FL
Univ Illinois, Dept Mech Engn MC 251, Gear Res Lab, 842 W;Taylor St, Chicago, IL 60607 USA;Univ Illinois, Dept Mech Engn MC 251, Gear Res Lab, Chicago, IL 60607 USA

15: MATSUO_K NAKAI_T KATO_C MORIYA_T ISODA_H TAKEHARA_Y SAKAHARA_H
Hamamatsu Univ, Sch Med, Dept Radiol, 3600 Handa, Hamamatsu,;Shizuoka 4313192, Japan;
Hamamatsu Univ, Sch Med, Dept Radiol, Hamamatsu, Shizuoka 4313192, Japan;
Minist Int Trade & Ind, Agcy Ind Sci & Technol, Electrotec Labs, Life Elect Res Ctr, Tsukuba, Ibaraki 305, Japan;
Toyohashi Sozo Coll, Dept Management Informat, Toyohashi, Aichi, Japan

16: ST GEORGE_M LICHTENBERGER_L BELLUGI_U
Salk Inst Biol Studies, Cognit Neurosci Lab, La Jolla, CA 92037 USA;
Univ Calif San Diego, San Diego, CA 92103 USA
Salk Inst Biol Studies, Cognit Neurosci Lab, 10010 N Torrey;Pines Rd, La Jolla, CA 92037 USA;

17: CLARK_MG ROSEN_GD TALLAL_P FITCH_RH
Univ Connecticut, Dept Psychol Behav Neurosci, 3107 Horse Barn;Hill Rd,U-154, Storrs, CT 06269 USA;
Univ Connecticut, Dept Psychol Behav Neurosci, Storrs, CT 06269 USA;
Rutgers State Univ, Piscataway, NJ 08855 USA;
Harvard Med Sch, Boston, MA USA

18: LIASIS_A TOWELL_A BOYD_S
Great Ormond St Hosp Children, London, England;Univ Westminster, London W1R 8AL, England

19: COPLAND_DA CHENERY_HJ MURDOCH_BE
Univ Queensland, Dept Speech Pathol & Audiol, Brisbane, Qld;4072, Australia;Univ Queensland, Dept Speech Pathol & Audiol, Brisbane, Qld 4072, Australia

20: DE KEIZER_NF ABU-HANNA_A ZWETSLOOT-SCHONK_JHM
Acad Med Ctr, Dept Med Informat, J2-256,POB 22660, NL-1100 DD;
Amsterdam, Netherlands;Acad Med Ctr, Dept Med Informat, NL-1100 DD Amsterdam, Netherlands;
Univ Utrecht, Sch Med, Julius Ctr Patient Oriented Res, Utrecht, Netherlands

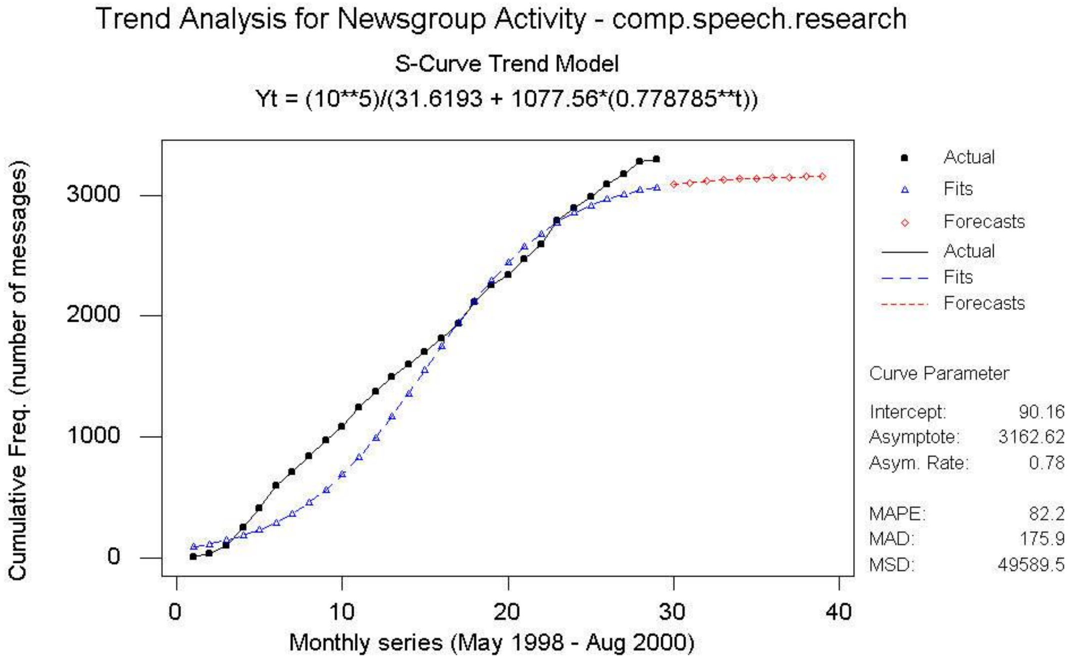
ANNEX VIII
Full list of Disciplines in which ELSnet researchers were granted a Doctorate

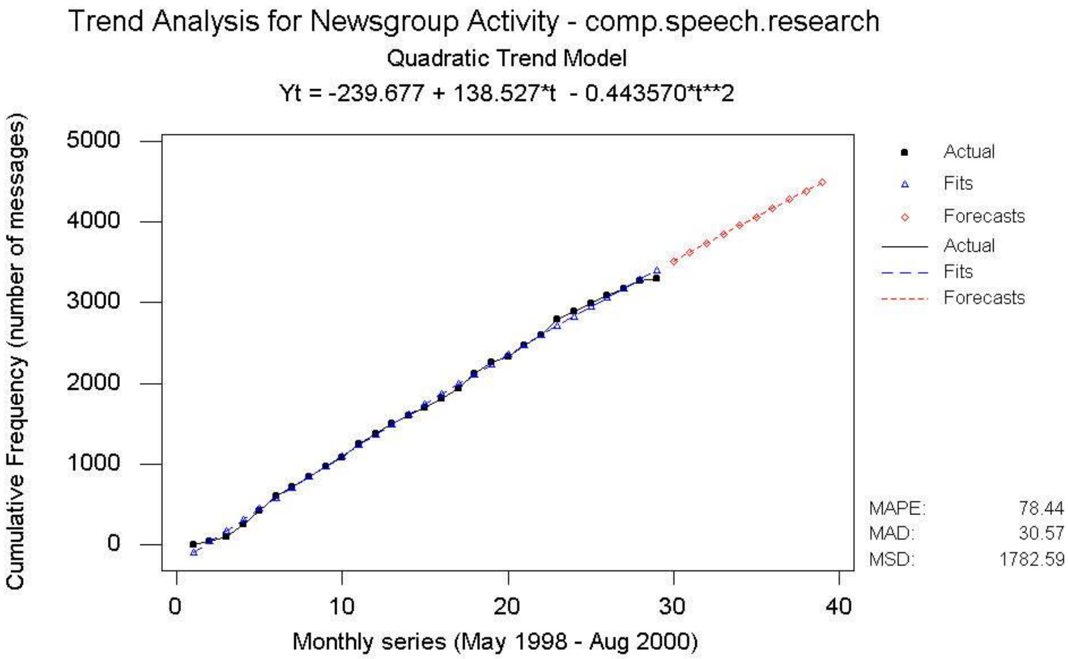
doctordiscipline

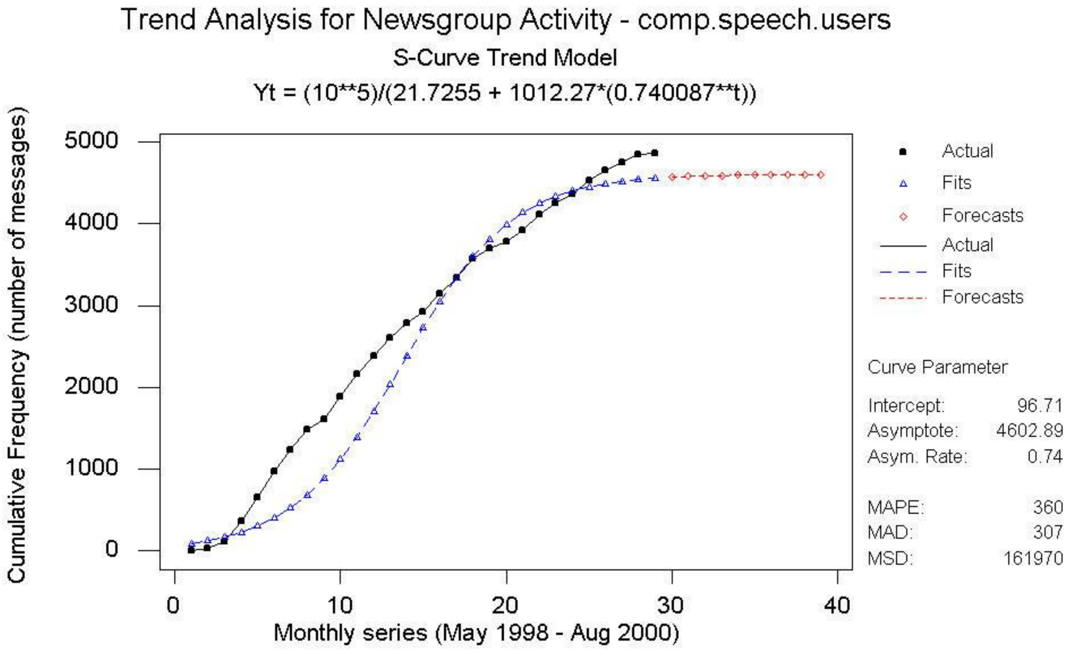
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	-	2	.6	.6	.6
		141	45.2	45.2	45.8
	Applied linguistics	1	.3	.3	46.2
	Artificial Intelligence - NL	1	.3	.3	46.5
	Artificial Intelligence	9	2.9	2.9	49.4
	Audotpry Processing an Rhythm	1	.3	.3	49.7
	Biochemstry & Signal Processing	2	.6	.6	50.3
	Cognitive & Linguistic Sciences	1	.3	.3	50.6
	Cognitive Sciences	3	1.0	1.0	51.6
	Computational Linguistics	13	4.2	4.2	55.8
	Computer Application	1	.3	.3	56.1
	Computer science	1	.3	.3	56.4
	Computer Science	28	9.0	9.0	65.4
	Computer Science / Electronics	1	.3	.3	65.7
	Computing	1	.3	.3	66.0
	Corpus Linguistics	1	.3	.3	66.3
	Dialogue Systems	1	.3	.3	66.7
	Digital Signal Processing	1	.3	.3	67.0
	Education	1	.3	.3	67.3
	EFL Methodology (grammar)	1	.3	.3	67.6
	Electrical Engineering	3	1.0	1.0	68.6
	Electrical Engineering & Computer Science	1	.3	.3	68.9
	Electrical Engineering / Speech Understanding	1	.3	.3	69.2
	Electronics / Speech	1	.3	.3	69.6
	English Language	1	.3	.3	69.9
	Faculty of Arts	1	.3	.3	70.2
	Folklore studies	1	.3	.3	70.5
	Historical Stylistics	1	.3	.3	70.8
	Human Communication Sciences	1	.3	.3	71.2
	Information Retrieval	1	.3	.3	71.5
	Information Science	1	.3	.3	71.8
	Information Theory	1	.3	.3	72.1
	Language and Speech	1	.3	.3	72.4
	Language Engineering	1	.3	.3	72.8
	Language Modelling / Speech Recognition	1	.3	.3	73.1
	Letters	1	.3	.3	73.4
	Linguistics	29	9.3	9.3	82.7
	Linguistics / Phonetics	3	1.0	1.0	83.7
	Mathematical linguistics	2	.6	.6	84.3
	Mathematical Linguistics	1	.3	.3	84.6
	Mathematical Logic	1	.3	.3	84.9
	Mathematical Sciences	1	.3	.3	85.3
	Mathematics	1	.3	.3	85.6
	Mathematics and Computer Science	1	.3	.3	85.9
	Medical Informatics	1	.3	.3	86.2
	Mother-infant communication, Ethology	1	.3	.3	86.5
	Music Acoustics	1	.3	.3	86.9
	Musical Acoustics	1	.3	.3	87.2
	Natural Language Processing	3	1.0	1.0	88.1
	Pattern Recognition	2	.6	.6	88.8
	Philosophy	2	.6	.6	89.4
	Philosophy/Cognitive Sciences	1	.3	.3	89.7
	Phonetics	2	.6	.6	90.4
	Phonology & Phonetics	1	.3	.3	90.7
	Physical Sciences	1	.3	.3	91.0
	Physics-Mathematics	1	.3	.3	91.3
	Physics	1	.3	.3	91.7
	Programming Language Research	1	.3	.3	92.0
	Psycholinguistics	1	.3	.3	92.3
	Psychology	1	.3	.3	92.6
	Romance Philology	1	.3	.3	92.9
	Room Acoustics	1	.3	.3	93.3
	Sciences	1	.3	.3	93.6
	Semiotic	1	.3	.3	93.9
	Signal Processing	1	.3	.3	94.2
	Signal Theory and Communications	1	.3	.3	94.6
	Slavic Languages and Literatures	1	.3	.3	94.9
	Speech Acoustics	1	.3	.3	95.2
	Speech Compression	1	.3	.3	95.5
	Speech Perception	1	.3	.3	95.8
	Speech Recognition	5	1.6	1.6	97.4
	Speech Science	1	.3	.3	97.8
	Speech Synthesis	1	.3	.3	98.1
	Speech Technology	3	1.0	1.0	99.0
	Statistiques Mathematiques	1	.3	.3	99.4
	Technical Cybernetics / Information Theory	1	.3	.3	99.7
	Theoretical Elementary Particle Physics	1	.3	.3	100.0
	Total	312	100.0	100.0	

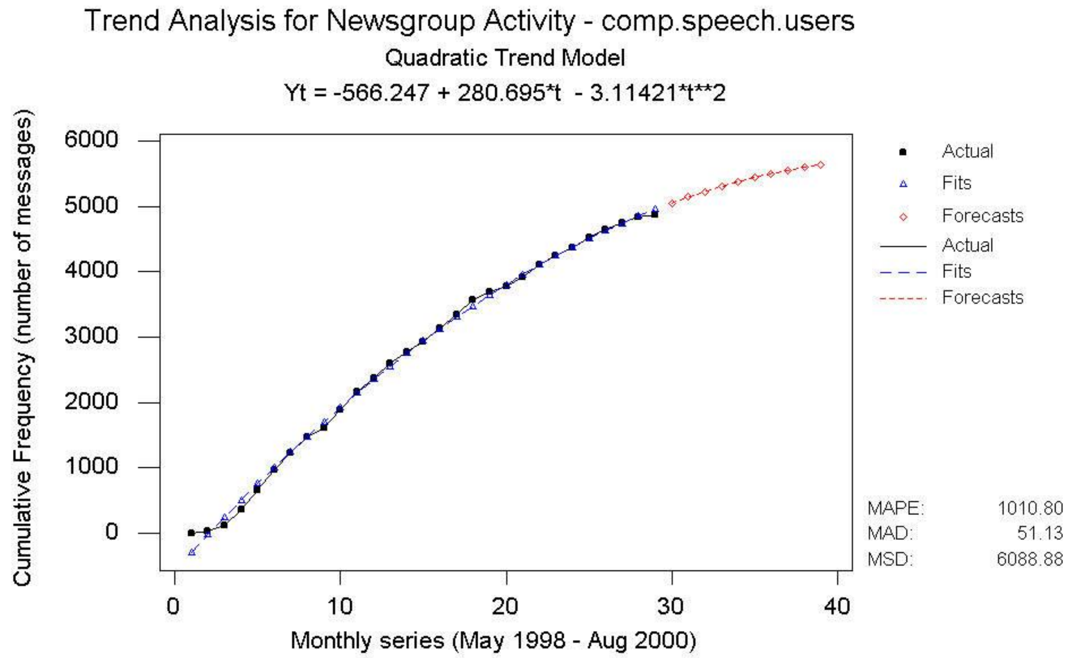
ANNEX IX

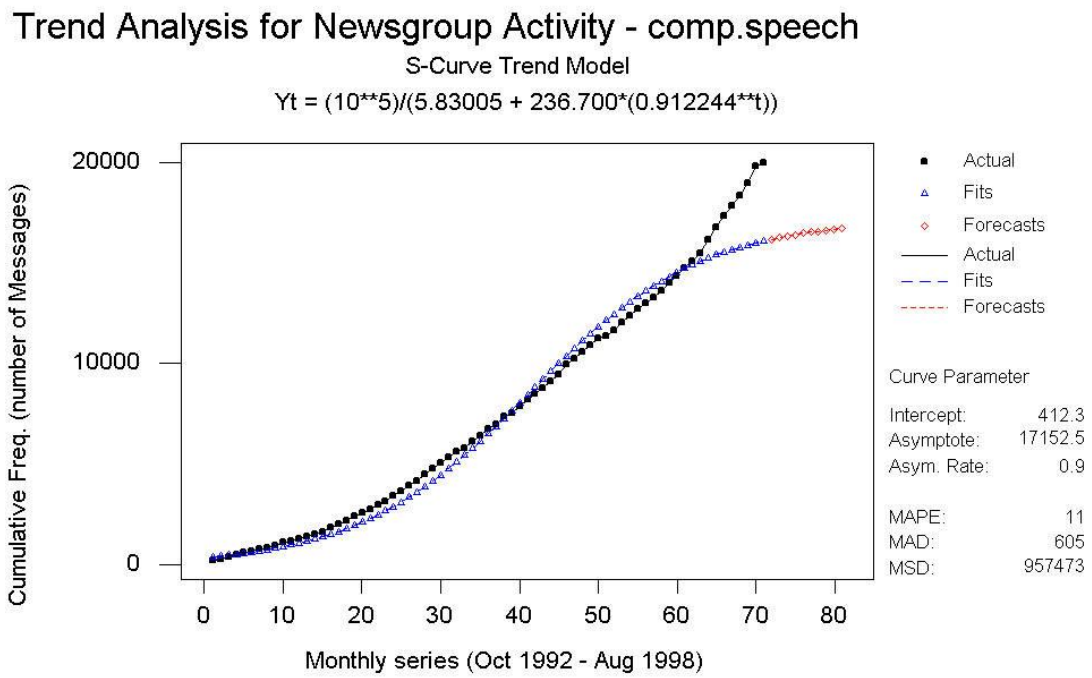
Growth Pattern of Speech Newsgroups (Quadratic and Logistic Trend Models)

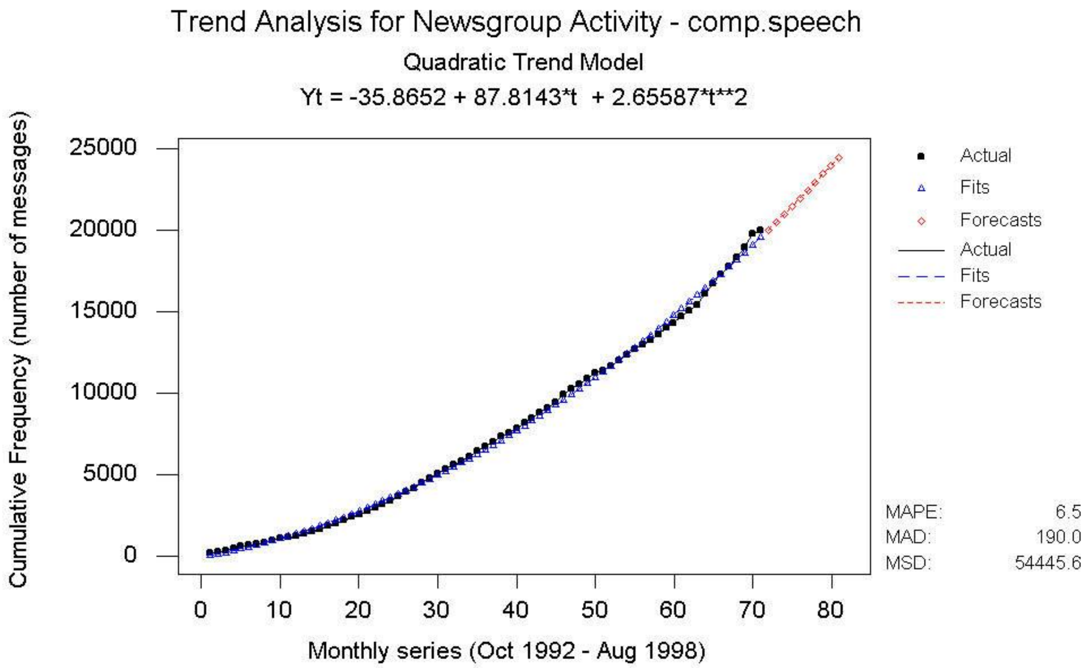





Trend Analysis for Newsgroup Activity - comp.speech.users
S-Curve Trend Model
Yt = (10**5)/(21.7255 + 1012.27*(0.740087**t))
Cumulative Frequency (number of messages)
Monthly series (May 1998 - Aug 2000)
Actual
Fits
Forecasts
Curve Parameter
Intercept: 96.71
Asymptote: 4602.89
Asym. Rate: 0.74
MAPE: 360
MAD: 307
MSD: 161970


Trend Analysis for Newsgroup Activity - comp.speech.users
Quadratic Trend Model
Yt = -566.247 + 280.695*t - 3.11421*t**2
Cumulative Frequency (number of messages)
Monthly series (May 1998 - Aug 2000)
Actual
Fits
Forecasts
MAPE: 1010.80
MAD: 51.13
MSD: 6088.88


Trend Analysis for Newsgroup Activity - comp.speech
S-Curve Trend Model
Yt = (10**5)/(5.83005 + 236.700*(0.912244**t))
Cumulative Freq. (number of Messages)
20000
10000
0
0 10 20 30 40 50 60 70 80
Monthly series (Oct 1992 - Aug 1998)
Actual
Fits
Forecasts
Actual
Fits
Forecasts
Curve Parameter
Intercept: 412.3
Asymptote: 17152.5
Asym. Rate: 0.9
MAPE: 11
MAD: 605
MSD: 957473


Trend Analysis for Newsgroup Activity - comp.speech
Quadratic Trend Model
Yt = -35.8652 + 87.8143*t + 2.65587*t**2
Cumulative Frequency (number of messages)
25000
20000
15000
10000
5000
0
0 10 20 30 40 50 60 70 80
Monthly series (Oct 1992 - Aug 1998)
Actual
Fits
Forecasts
Actual
Fits
Forecasts
MAPE: 6.5
MAD: 190.0
MSD: 54445.6

ANNEX X

Questionnaire implemented in the Electronic Survey to ELSnet network researchers

The electronic questionnaire implemented to ELSnet network members was made available at the following URL:

http://www.sussex.ac.uk/Users/prpb7/speech/survey/ELSnetsurvey.htm

An hard-copy of the electronic questionnaire is presented in the following 13 pages of this Annex.