Chapter I.

Studying International Collaboration in Science: Research Questions and Methodologies

Collaboration is a way to expand my horizon... Learning something new is a sign of its success. S. Shelah, a mathematician interviewed for the project
Introduction

International collaboration in science has been growing at a significant rate. From a low base in the mid-20th century, when collaboration was largely institutional, local, or within national borders (Price 1963; Beaver 2001), international collaboration now makes up a considerable share of scientific research. When counted by co-authorships, internationally co-authored articles accounted for 15 percent of all published articles in 1998, up from less than 8 percent in the 1980s. (NSB 2000) Figure 1 shows the increases in internationally co-authored articles for 10 scientifically advanced countries over a 10 year period.

![Graph showing international collaboration growth](image)

Figure 1. Increase in the percentage of national articles that are internationally co-authored, scientifically advanced countries, 1986-1997
Data: NSB 2000

The rapid shift in the organization of scientific research to include a significant amount of international collaboration is itself a good reason to undertake a study. The question is made more interesting by a number of factors that would seem to militate against a growth in geographically dispersed collaboration. These include language and cultural differences, different institutional incentives and rewards, and different national missions guiding research fund allocations. As this thesis will show, no explanation has yet been offered that can fully explain the rapid rise in international collaboration. As a result, no reliable account of the implications of these changes for governance of science has been detailed. This thesis attempts to bridge this gap in the literature by offering a theory of the dynamics of knowledge creation within international collaboration in science.
This introduction has five sections designed to introduce the research that is presented in the following chapters. The five sections presented in this introduction are: 1) an overview of the study of collaboration with a brief literature review and research questions; 2) expectations held at the beginning of the research, and 3) definitions and scope in this thesis: 4) a description of methodologies; and 5) the organization of the thesis.

1. The study of collaboration

International collaboration in science is a geographically dispersed form of collaboration. Since Price (1963) began a discussion about the increasingly collaborative nature of science, there have been a number of studies from bibliometric, sociological, and political perspectives examining the phenomenon. From the bibliometrics perspective, the study of collaboration has often taken the shape of counting co-authorships (Price 1963; Carpenter & Narin 1973; Lindsey 1980; Narin 1991; Luukkonen et al. 1993; Miquel 1994; Doré 1996; Georgiou 1998; Glänzel 2001; Melin & Persson 2000; Laudel 2002). From a social perspective, Beaver & Rosen (1978a and b) examined the professionalism of science as a motivating factor for collaboration. Gibbons et al. (1994) suggested the role of transdisciplinarity as a driving factor. From a political or policy perspective, Crawford et al. (1993) and Skolnikoff (2001) and examined the political role of scientific cooperation, while Katz & Martin (1997) and Smith & Katz (2000) have looked at the policy structures supporting cooperative research.

Of particular relevance to the research presented here are three subsets of science studies. One takes an observational approach, examining the actions of scientists to see what it is they actually do when they collaborate. This line of research has included the work of Latour (1987), Knorr-Cetina (1999), Laudel (2001) studying the communications and personal dynamics in collaborations. They often focus on geographically proximate (often side-by-side) collaborations. Others have approached collaboration from the perspective of communications media as they affect the willingness and ability of researchers to work across geographical distances (Siegel et al. 1986; Starr 1995) focusing on the role of the Internet. Meyerowitz (1985) examines the influence of communications media on the structure of collaboration itself, a theme taken up by and discussed by Ziman (1994), and also by Gibbons et al. (1994) as part of their description of Mode 2 research. Still others have examined the reward structures within science, at the institutional and
field levels, to see how these features influence the willingness of scientists to collaborate. (Merton 1973; Whitely 1984)

Recently there has been attention on the increasingly international character of collaboration from a quantitative perspective. These studies have examined collaboration at the country-to-country level (Okubo et al. 1992; Zitt et al. 2000; Glänzel 2001); patterns of relationships among researchers within disciplines of science (Frame & Carpenter 1979; Schubert et al. 1989; Glänzel 2003); and at the global level (van Raan 1997; Georgiou 1998; Luukkonen et al. 1992). A more qualitative and politically-oriented set of inquiries can be found in Crawford (1992) and Crawford et al. (1993), suggesting that international collaboration should in fact be called "transnational" because it appears to transcend the national interests of the participants. In a comment that anticipates the research presented in this thesis, Crawford (1992) writes: "The work of mapping these [international linkages] historically, which has hardly begun, may well call for a fourth category: networks and other forms of informal and transitory groupings." (p. 21)

In a separate stream of literature, and even more recently, network theorists have taken an interest in scientific collaboration, largely because of the readily available dataset created by scientific co-authorships and citations in databases that facilitate network analysis (Albert & Barabási 2000; Barabási et al. 2002; and Newman 2001, 2004). These scholars approach the data by examining collaborations at the level of individual scientists, not from the international or social perspectives, to demonstrate and examine the dynamics of interactions within large networks. The interesting and useful part of these inquiries is the not just the creation of methodologies to explore the dynamics of networks, but the ability within social science to apply this analysis to communications systems. (Otte & Rousseau 2002)

These previous inquiries and approaches, combined with the data on collaboration, prompt a fresh look at the following questions: Given that international collaboration is growing at a significant rate, why is this happening? What does this mean for the dynamics of knowledge creation in science? What does the shift in organization mean for governance of science? What influence will the rise of international collaboration have for the exploitation and dissemination of research results? These became the research questions for this thesis.
2. Expectations of research outcomes

The research project reported here began with a set of expectations that guided the research and influenced the choice of methods. The list of expectations is reproduced here from the proposal written at the beginning of this project. (June 2002) A reflection on how the project supported or refuted these expectations is discussed in Chapter VIII.

- **International collaboration will be found to be increasingly conducted in a distributed research mode.** This expectation grew out of my RAND research that preceded the research at the University of Amsterdam. The RAND research showed that an increasing number of research projects being funded by the U.S. government were investigator-driven linkages. Although “megascience” projects were still being funded, the growth in funding for international projects appeared to be in geographically dispersed, scientist-initiated research. Moreover, a review of existing literature shows that the fields in which international collaborations are growing are fields where the subject of research is not dependent upon large-scale equipment, but are motivated by other factors. In addition, this expectation grows out of my RAND analysis of S&T capacity that found that, by 1998, as many as 60 countries can be considered at least “scientifically proficient,” possibly offering an increased number of scientists available for international collaborative projects. Research talent may be more widely distributed around the world, perhaps facilitating the trend towards distributed research.

- **The sub-fields of science that constitute growing areas of international collaboration are self-organizing ("bottom-up") and scientist-initiated, not in fields where research is initiated by governments or large international organizations.** They will have a network structure exhibiting dynamic features. The projects that attracted government funding in the 1950s through the 1980s, in the majority, were large-scale, equipment-based research. These projects served national political goals and were hierarchical ("top-down") in their organization. Since the 1980s, it appears that international collaboration is more often initiated by the individual scientist—not necessarily tied to a megascience project—and that, in some cases, governments have been persuaded by the scientific community to fund distributed research projects. It is not clear why this is occurring, but I will be exploring reasons for this shift in this research project. The expectation is that international collaboration in science is more often self-organizing—initiated by researchers themselves and
conducted across geographical distances and that it can be characterized as a network. The network of interactions is expected to have an internal dynamic different from the dynamics of other systems.

- **The communications that constitute dynamic, growing research area are more often initiated through face-to-face communication rather than through the Internet.** Although some groups have suggested that the Internet and other communications tools are a causative factor in the growth of collaboration, research in interpersonal dynamics, and the results of RAND surveys of how scientists initiated collaboration, suggest that face-to-face communication is more important to identifying potential collaborators than other forms of communication tools. The expectation is that scientists will initiate international collaborations more often through face-to-face communications than through other media. This question will be explored principally through interviews.

- **Once a communication is initiated between scientists, a number of information technology tools are used to exchange information.** Distributed communication requires that researchers actively communicate in real time. As a result, my expectation is that the enabling features of the Internet, facsimile machines, telephony, and HTML will all be factors contributing to the dynamism of distributed communication. This type of real-time, remote communication requires new ways of conducting research—methods that are changing the way scientific research projects are managed.

- **Practitioners from more than one traditional scientific discipline will be involved in distributed research projects.** Gibbons et al. (1994) suggest that the collectivised, team-based approach to R&D is a new method of knowledge production, called Mode 2, resulting from the expansion in the supply of knowledge producers and the demand for specialised knowledge. I expect that some features of the Mode 2 theory of knowledge creation will be upheld by a closer examination of international collaboration.

- **Funding for collaborations will come from at least two sources and will not be a major motivation for international collaboration.** While I expect that communications within international collaboration in science will have self-organizing features, science remains an activity largely funded by governments. Governments are reluctant to provide funding directly to foreign researchers. In
order for a research project to proceed, it is probable that both researchers will need to find funding. As a result, in interviews with researchers, I expect that they will report that each researcher in the collaboration brought some funding or in-kind contribution to the project.

- Fewer than 10 percent of the collaborations studied will have participants from developing countries. Developing countries generally do not have a great deal of public funds available for research and development. As noted above, international collaborations often require that each researcher fund their own participation. As a result, I expect to find that very few international collaborations include researchers from developing countries. This is not to say that researchers in these countries are not "world-class," rather that they will find it difficult to participate in distributed research because of some of the features I have discussed above.

- The research subjects that characterize international collaborations will be in rapidly growing fields, but will not represent highly specialized (or "niche") research topics. Existing literature reports that international collaboration appears to be growing in many fields of science. If specialization were the driving force behind collaboration, then one might expect to see a clustering of researchers collaborating in some fields, and other fields where there is little collaboration. I expect to find that other reasons—possibly within the communications system providing the basis for knowledge creation process itself—are motivating collaboration.

3. A social science view of the natural sciences: Defining terms

Despite many articles and books written about knowledge creation and the organisation of science, few actually define science; they seem to presume a common understanding. Even so, the edges of this understanding are not clear, and perhaps are not held in common. Originally derived from the Latin, scientia, knowledge, over centuries, the term science has developed in both historical understanding and philosophical discourse. Aristotle argued that science is a method, and that the subject of scientific inquiry existed external to the viewer. (Metaphysics) His understanding of science, however, suggests that he considered all knowledge to be contained within science, rather than science within knowledge. Yet, the latter would be more consonant with current views.
For the purpose of this thesis, science as a widely accepted, adoptable, and transferable set of methods and assumptions about how to understand the world. As scientific practices become standardized and transferred, the practice of science becomes a system of knowing things. That system takes on a dynamic that transcends and subsumes the practitioner. As a result, science is often referred to as a "universal" way of knowing because the knowledge passes without regard to the political allegiances, gender, race, or other attributes of the person seeking knowledge. The universal nature of science can be demonstrated by the gradual expansion of the scientific enterprise over the past 300 years from a few small centers in Europe to hundreds of centers around the world. (Ben-David 1971)

By scientific research, I mean the socially-based activities that seek knowledge about the natural world through observation and experimentation, generating a disciplinary and multi-disciplinary body of knowledge. This thesis bounds science as a set of verifiable information that cannot be changed by human action. Thus the study is limited to what are sometimes called the "hard sciences." The results of these inquiries can be understood and reproduced by other knowledgeable practitioners regardless of their social status, political allegiances or geographic location. As such, the knowledge created and codified as a result of these activities has a "universal" aspect in that any knowledgeable and interested person can understand and use the information. A discipline within science, to extrapolate from Kuhn (1962), is the community of trained specialists investigating a well-defined range of problems with methods and tools specifically adapted to the task. The definition of the scientific problems and resulting methods of investigation in Kuhnian science derive from a professional tradition of theories, techniques, and skills acquired through prolonged training and acculturation. (Kuhn 1962)

Collaboration refers to those activities where scientists work together on a specific research project or question. (Laudel 2001; Katz & Martin 1997) It involves parallel or sequential sets of scientific observations or experiments towards a common goal. Cooperation is more broadly conceptualized here as the sharing of ideas and concepts among colleagues and other interested parties, similar to the exchange of information that has been a feature of science for two centuries. (Beaver & Rosen 1978a) From the 17th through the 19th century, science largely was undertaken by a single researcher who then reported his (or, more rarely "her") findings to an elite community. Derek de Solla Price (1963) calls this style of research "small science"; Gibbons et al.(1994) call it "Mode 1"
research. The sharing of information among these practitioners constitutes cooperation in science.

Collaboration is understood within this thesis as a more specific activity. It involves practitioners working together on a scientific problem or question towards a common research goal. It is a specific, time-limited interaction measured by co-writing of grant proposals, sharing problem definition and experimentation, and co-authoring of peer reviewed literature. Price (1963) was among the first to point out that an increasing number of scientific papers were being co-authored, suggesting that science was becoming more collaborative. Moreover, he noted that the entrance of government funding into the scientific enterprise, the tying of science to economic growth, and the increasing number of students trained in the sciences, had expanded the context within which science takes place, ushering in an era of “Big Science.” Price shows that, as “Big Science” pumped funds into the system, collaboration was also increasing geometrically, although he does not claim that the two features of modern science are necessarily linked. These distinctions will be explored in more detail in later chapters.

4. Methodologies for exploring knowledge creation within international science

The study described in this thesis has employed many methods, both qualitative and quantitative, to uncover and expose the dynamics of knowledge creation in global science. This section describes the different methods used, focusing on the quantitative and interview methods. The research proceeded from the broad to the specific: moving from an analysis of the global network of science to networks within disciplines of science, and then to the methods and motives of communication among scientists themselves. This chapter takes a similar approach, beginning with a description of the network analysis conducted at the global level, then analysis conducted at the disciplinary level. The interview methodology is then discussed.

The techniques described here were chosen for their ability to reveal the dynamics of interrelationships among scientific researchers at the global level. Systems and networks that contain knowledge cannot be considered as “given” or immediately available for observation or measurement. (Biggerio 2001; Leydesdorff 2001(a)) The analyst has to specify the relevant system before it can be indicated or measured. As discussed above, and following Hesse (1974 and 1980) and Leydesdorff (2001(a)), I expect that science at
the global level is a communications system with a network structure exhibiting dynamic features.

Dynamics within any set of relationships (such as a within a social network) are difficult to measure. The difficulty of studying dynamics of global science derives from several of its features: Dynamics are evidenced by communications, which are often ephemeral, and occur at many levels of formality.\(^1\) They can only be identified by the "traces" they leave behind (Leydesdorff & Scharnhorst 2003), and then, only the codified (not tacit or informal) communications can be quantified. By studying the traces (in this case, co-authored articles and citing relationships) indicators of networks can be created. Measurement is also made difficult in part because, in order to study dynamics, time must be introduced as a dimension. (Leydesdorff, 2001(a))

The relevant system being examined here is the international system of science, or the system of scientific communications that operate at the global level. While these communications go on between people, the indicators that are available to evidence the relationships at the global level are based on the addresses linked to countries where these authors claim to be located. Using countries as part of the indicators for dynamics at the global level also allows me to test for the growth of geographically dispersed research. In addition, I am interested in the communications tools that are being used (print, Internet, databases) to see if information and communication technologies are having an effect on the dynamics of knowledge creation.

In addition to the study of global linkages over time, we must add other layers of analysis. Science is difficult to study in part because communication processes are multi-layered. (Leydesdorff 2001(a)) As a result, the techniques used here are conducted in a cascading or "nested" fashion in an effort to gain different views into the dynamics of the relationships at different layers and over time. The layers examined are the meta-network at the global level, disciplinary networks at the global level, citing relationships at the disciplinary level, and individual communications among scientists in terms of patterns of attachment, as well as how scientists appreciate these communications reflexively. These different layers combine to provide different insights into communications within the

---

\(^1\) Communication within self-organising systems is distinguished in three ways:
- Degree of formality (formal or written/codified vs. informal or spoken/tacit)
- Degree of richness (from face-to-face to virtual) or (phone, fax, regular and electronic mail)
- Size (single message, dyadic, or collective). (Biggiero 2001)
network of global science. At the end of this chapter, we will discuss the strengths and weaknesses of these approaches as they apply to the study.

The techniques described here are used to study scientific relationships both as a static "snapshot" in time and across time. The 10-year interval used in this study, between 1990 and 2000, provides some insight into the changing communications patterns over time. In addition, this decade is a particularly interesting time to study. Five events significantly changed the environment for science during this time: 1) the break-up of the Soviet Union and the end of the bi-polar world, 2) the reunification of Germany, 3) the growth of information and communications technologies (specifically, the Internet and the World Wide Web), 4) the rise of the European Union and the growth in its funding for science, and 5) the globalization of industry. Each of these factors has influenced either the supply of scientific knowledge (the newly independent states have been encouraged to expand their participation in global science, for example) or the demand for science (the local information requirements of global businesses have encouraged investment in research). Each of these factors has also affected the political process that provides funding and support for science.

4.1 Mapping the network of international science

The place to begin examining global science is by detailing the patterns of linkages among researchers from different nations. Both quantitative and qualitative approaches can shed light on this question, and both are used in this study. The first part of the analysis is a quantitative study of the networks evidenced by co-authorships at the international level. This measured the co-occurrence of authorships among practitioners in different countries in the two years examined.

Figure 2 illustrates the method used to map communications among researchers at the international level. The figure illustrates the steps taken, processes used, and products created. The analyses included data identification, database creation and management, and matrix management and analysis in steps 1 through 4. Between steps 3 and 4, I considered different counting schemes that could be used to examine the relationships that were being revealed in the matrices. (The choice of integer counting is

---

2 There are a number of different descriptions of the process of globalization of industry. One useful description, directly relevant to this paper, is contained in the Science & Engineering Indicators 2000. (NSB, 2000).
described in detail below.) Then, once the co-occurrence tables were created, in step 5, I considered different ways to expose the relationships that were implicit in the matrices. These choices, based in part on practices detailed in literature and in part on the application of new tools developed for this research, allowed both bilateral measures and multilateral measures. I chose cosine analysis, factor analysis, and affiliations as the most illustrative for the purposes of my study. The application of these analytic tools allowed different types of outcomes. Steps 6 and 7 involved applying the tools to the data and revealing and illustrating the networks. These choices are also detailed below.

Figure 2. Steps in the process of conducting co-author analysis

The concepts behind this approach are drawn from bibliometric and social network analysis (SNA). The bibliometric approaches involved counting links created by studying author addresses and by detailing the citing environments of journals. These tools were primarily developed by Leydesdorff (2001a), and Leydesdorff & Cozzens (1993), Luukonen et al. (1993), and Glänzel (1990). These are standard tools and methods for assessing the structure of scientific communications.

Once it is collected from the Science Citation Index CD-Rom version, the bibliometric data was analysed using network analysis. Several authors have suggested that science can be conceptualized as a network structure. (Hesse 1974 and 1980; Callon & Law 1989; Leydesdorff 2001(a)) Leydesdorff (2001(a)) has noted that relations among
texts can “provide us with a network model of scientific developments.” (p. 57) Derek de Solla Price (1965) suggested that co-authorship of journal articles reflects an underlying network. Relatedly, approaching science as a social network has been a tool of social science analysis (Wellman & Berkowitz 1988) but it has not been as widely applied to science studies.

Social network analysis (SNA) is not a formal theory in itself; rather it is a strategy or a tool for investigating and exposing social structures. (Scott 2000; Otte & Rousseau 2002) SNA began to be widely used in the early 1980s and it has dispersed to a wide number of fields over time. (Otte & Rousseau 2002) There has been an increasing interest in using SNA in recent years, perhaps because of the increase in desktop computing power, databases, and software tools which facilitate the application of this type of analysis to many different fields. Author co-citation has been explored as a measure of networks within science at least since White & Griffith documented the capacity in 1981. (White & Griffith 1981)

In SNA, and in the research supporting this thesis, for the purposes of exposing the network itself, the motivations of actors are not the primary focus: the relationship or formal, codified communications between actors is the object of study. As Leydesdorff and Scharnhorst have noted, “Organized knowledge production can…be considered as the codification of communication. Communications leave traces that can be studied as indicators.” (2003, P. iv) Relational data in the form of addresses are the subject of investigation. Citations and co-citations are another way to trace knowledge networks, but these indicators are not directly linked to collaboration, and therefore are not a subject of this study.

4.2 A brief description of graphs and network theory

The findings of this research depend on understanding graph theory and network theory, so these are introduced briefly here. Social network analysis allowed me to explore and visibly depict the intensity and dynamics of interrelationships among researchers from different countries at the international level. Researchers have found that many real-world systems take on the form of networks. (Jin et al., 2001) In research similar to that undertaken for this thesis, Newman (2000) defines a social network as a “collection of people, each of whom is acquainted with some subset of the others.” Such a network can be represented in a graph as a set of points or nodes (vertices) denoting actors
(people or countries), joined in pairs by lines (edges or arcs) denoting acquaintance. or, in the case of this study, co-authorships. Vertices that attract many links are called "hubs."

In social network analysis, social structures are formally represented as graphs, and structural properties of these graphs are assumed to be useful in the explanation of social phenomena.

Large-scale networks that have been studied include the World-Wide web (Huberman 2001), transportation networks such as air routes, blood vessels, and the telephone system. In the past few years, there has been a substantial amount of work done on network structure and function; much of this has emerged from within the physics and biological sciences communities. (See referenced works of Strogatz, Watts, and Newman). The most interesting of this research for our purposes is the attention that has been given to networks that are self-organizing and evolving over time.

Using the SCI CD-Rom 1990 and 2000 data, I sought to apply social network analysis to examine the intensity and dynamics of interrelationships among researchers from different countries. The population of "nodes" that comprise the social network are scientists who have co-authored papers with colleagues from other countries. The data are arranged in an N x N matrix, which contains one row and one column, with a number indicating the occurrence of a co-authorship event between practitioners.

Social network analysis derives from a matrix where each node within the network is listed on both axes, and the resulting cells indicate the presence or absence of links from each node to each other node within the network. For example, in Figure 3 the 1's in the matrix on the left (known as an "adjacency" matrix) provide a representation of the links in the diagram on the right (known as a "sociogram").

![Figure 3. An illustration of a sociogram](image-url)
With such a small number of nodes, it is easy to see some of the basic properties of the network. For example, node B occupies a distinctive and relatively "central" position within the network with more ties than any other node. Nodes A, B, and E form a tightly connected "clique" in which every node is tied to every other node. The same is true of nodes B, C, and D. Node B provides a bridge between these two cliques, indirectly linking otherwise unconnected nodes. A quick count reveals a total number of six links out of a possible maximum of ten (the number of links that would have existed if every node had been directly linked to every other node). This equals 60 percent of the possible links (a percentage referred to as the "density" of the network). With a larger number of nodes, it is impossible to identify the structural properties of the network simply through observation. Instead, it is necessary to manipulate the underlying matrix with specialized software to reveal such properties as the overall density of links, the division of the network into relatively distinct cliques or clusters, the degree of connectedness or centrality of specific nodes, and so forth. The most widely used software for these purposes is Ucinet. (Borgatti et al. 2002)

For large networks, the interrelationships among the nodes of the network are not clear or easily determined. Ucinet provides the algorithms that find the interrelationships and affiliations among networks. Bringing the matrix of co-occurrences/co-authorships into Ucinet for analysis, and then into Pajek for the visualization of the social networks, I sought to identify the clusters representing collaborations.

4.3 Examining Co-authorships

An effective way to quantify communication linkages is by using co-authorships. A scientific document is co-authored if it has more than one author, and it is internationally co-authored if at least two of these authors list addresses in different countries. Data on co-authored articles can be retrieved from almost any bibliographic database. However, the Science Citation Index produced by the Institute for Scientific Information is the most reliable source for a comprehensive survey. I drew data from the Science Citation Index CD-Rom from 1990 and 2000 to create global networks for these years. The addresses

\[3\] The CD-Rom version is preferable to the SCI Extended version found on the Internet. The Extended version is regularly updated, and therefore, it would be very difficult to repeat this analysis, even from day to day. Assuming that the extended index can change, the data set could be altered and therefore the results presented here could not be reproduced.
make it possible to study co-authorships using nations as the identifying feature. The research did not distinguish among types of contributions (reviews, letters, proceedings, and journal articles) because I was seeking social connection to reveal the structure of the network regardless of the scientific status of the output.

When using co-authorships as the indicator of international linkages, there are a number of validity issues to consider. Scientific collaboration may lead to a range of outcomes, and a co-authored article is only one of these outputs. Katz and Martin (1997) have detailed a number of issues associated with using co-authorship data as an indicator of collaboration. They note that collaboration does not necessarily lead to co-authored publications. Conversely, co-authorship does not necessarily indicate collaboration: for example, research leaders may list their names on articles without actually contributing to the work. In many cases, although international collaboration is taking place, co-authors list a common institution in which they are working, thus leading to an understating of international linkages. A level of uncertainty is inherent in this data. As Melin and Persson (1996) have suggested, we can hope that “significant scientific collaboration leads to co-authored papers in most cases, the main reason being the priority claims of the scientists involved.”

While it would have been possible to examine the networks using other features of collaboration, these would not have served the analysis as well. For example, it would be possible to use the names of institutions as the unit of analysis, and examine links between institutions. This is a method of analysis commonly used, for example in the Triple Helix approach to examining dynamics of knowledge creation. (Leydesdorff & Etzkowitz, 2001) However, this would place too much emphasis on the physical location of the research rather than the dynamics of communications. It is the latter question that is explored in this research, not the physical place where it occurs. Similarly, it would have been possible to focus the analysis at the level of a city or geographic region rather than at the level of the nation. This would also provide interesting results. Nevertheless, this thesis takes the approach of asking the question of how international cooperation in science is affecting dynamics of science. To a great extent, science is still funded through a political

---

4 The United Kingdom is considered here in its component parts. Addresses are provided as England, Scotland, Wales, and Northern Ireland, and each is handled as a separate political unit for the purposes of this analysis.
governance structure. This compels the use of the nation as the point of analysis, rather than looking at other interesting links among regions, cities, or institutions.

Counting and representing the patterns of linkages among scientists can be done in several ways. I agree with the finding of Luukkonen et al. (1993) that the "choice of measures for gauging the strength of the ... relationships clearly affects the findings." (p. 33) Accordingly, I examined the SCI data using different methods. Although I used other methods to analyse the data, the findings presented here represent my assessment, conducted with Loet Leydesdorff, of the most illustrative ways to analyse and visualize the data when one is seeking to study network dynamics. I concur with Luukkonen et al., (1993) who suggest that it is important to use both absolute and relative measures if one wants to acquire a full picture of collaborative linkages. Absolute measures allowed me to see the overall size of the network, as well as which countries are in the center and which are at the periphery of the network.

Relative measures normalize size and therefore provide insight into the intensity of links between countries. I also applied these measures in an effort to reveal the architecture of the network. Relatedness measures such as the Salton Index (i.e., the cosine) illuminate the architecture of the relationship of countries. The Salton Index can be applied using social network analysis, showing intensity of relationships. Multilateral measures take the global network into account based on the assumption that the expected numbers of linkages between countries occur in proportion to a country's share in the global network.

Factor analysis allows the data to be rotated to find the correlations and find commonalities and differences among countries that are not obvious from a visual inspection of the data, and which do not become apparent using (hierarchical) relatedness measures. Factor analysis presents clusters of countries with strong relationships as well as intensities of relationships visualized in a multi-dimensional space.

I use these three sets of measures: descriptive statistics of observed values, relatedness measures (illustrated through social network analysis), and factor analysis to assess the absolute and relative positions of countries in the global system. They allowed me to gain a "snapshot" of what the global science network looked like in the two years studied. The ten-year interval between 1990 and 2000 allowed me to compare the two years to explore the changes in the network over time.

Steps in the Process

16
The process of examining co-authorships to identify patterns and networks of cooperation had 7 steps. These are illustrated in Figure 2. The process began by collecting into a single data set all papers produced for the relevant year. Table 1 inventories the data set at the global level; this data is discussed in detail in Chapter V.
The first analytic task is determining how to attribute papers to participating countries. Previous efforts to count co-authorships have used fractional, links, and integer counting in order to compare performance of countries or to examine networks. Fractional counting attributes the numbers proportionally, so that the number of authors on any given paper reduces the share of each participating country. A second way of counting is to identify the number of links represented among the countries involved, with each bilateral relationship counting as "1." The normalization in terms of number of links is more common in network analysis (Newman, 2000). A third way of counting is integer (Luukkonen et al. 1993) or whole/distinct count (Zitt et al., 2000) that attributes a count of "1" to each occurrence of authorship by a country created by the participation of researchers from that country. Table 2 shows how the different methods of counting affect the totals. Since I am interested in using counts to see how the network is developing between countries, I used integer counting.

During the counting phase, I applied a two-tiered analysis. The data was placed into an asymmetrical matrix containing raw occurrence data that recorded all countries across one axis, and all articles on the other. A count is placed in the corresponding cell created by this matrix. Table 3 provides an illustration. The digit in the cell represents the number of unique addresses for each country appearing in the address line of the respective article. This occurrence table was used to conduct factor analysis in SPSS to identify patterns within the global network.

<table>
<thead>
<tr>
<th>Year</th>
<th>Unique documents in SCI</th>
<th>Addresses in the file</th>
<th>Authors for all records</th>
<th>Internationally co-authored records</th>
<th>Addresses, internationally co-authored records</th>
<th>Percent internationally co-authored documents</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>778,446</td>
<td>1,432,401</td>
<td>3,060,436</td>
<td>121,432</td>
<td>398,503</td>
<td>15.6</td>
</tr>
<tr>
<td>1990</td>
<td>590,841</td>
<td>908,783</td>
<td>1,866,821</td>
<td>51,596</td>
<td>147,411</td>
<td>8.7</td>
</tr>
</tbody>
</table>

*Table 1. Data used to create international co-authorship tables*
Table 2. An example of different methods of counting co-authorships

<table>
<thead>
<tr>
<th>Method</th>
<th>USA</th>
<th>Switz</th>
<th>Scotland</th>
<th>Germany</th>
<th>France (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Integer</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Fractional</td>
<td>0.16</td>
<td>0.16</td>
<td>0.16</td>
<td>0.16</td>
<td>0.32</td>
</tr>
<tr>
<td>Links</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>8</td>
</tr>
</tbody>
</table>

Table 3. A sample of data in an asymmetrical matrix of the occurrence of authorship in an internationally co-authored paper (SCI 1990)

I then converted the occurrence table into a binary matrix of only ones and zeros ("integer counting") in order to construct a symmetrical matrix of countries appearing on both axes, with the co-occurrence of addresses appearing in the corresponding cell. Table 4 provides an illustration of the aggregated numbers that result when the names of countries appear on both axes. (The co-occurrence table was also used to conduct social network analysis, and is the representation of the sociogram notionally illustrated in Figure 3.)
Table 4. A sample of data in a symmetrical matrix of co-occurrences of authorship (a "sociogram") between countries in internationally co-authored papers, 1990

Once the data were collected into the co-occurrence matrix for 1990 and 2000, two types of analyses were applied: bilateral similarity and social network analysis, and multilateral similarity analysis (See also Figure 2, for steps in the process). Previous analysts have used a number of different approaches depending upon the question they were bringing to the data. Two measures have been used in scientometric analyses to achieve the weighted results (Luukkonen et al., 1993): 1) bilateral similarity measures, and 2) multilateral (pattern) similarity measures (e.g., Pearson correlations). The first method – bilateral similarity measures – I also found very helpful for these purposes.
The bilateral measures weight the data in order to illustrate links between separate pairs of countries. The resulting figures are derived from the observed number of shared papers \( C_{xy} \) of the countries x and y, weighted by the total number of papers, \( C_x \) and \( C_y \). Table 5 shows these measures and equation.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Salton’s measure</td>
<td>( S_{xy} = \frac{C_{xy}}{\sqrt{C_x \times C_y}} )</td>
</tr>
<tr>
<td>Jaccard’s measure</td>
<td>( J_{xy} = \frac{C_{xy}}{(C_x + C_y - C_{xy})} ) or ( \frac{\cap C_y}{C_x \cup C_y} )</td>
</tr>
</tbody>
</table>

*Table 5. Relatedness measures and equations considered to analyse the data*

Luukkonen *et al.* (1993) state that "The Jaccard measure underestimates the collaboration of smaller countries with larger ones, but the Salton measure underestimates the collaboration of smaller countries with each other." In a careful analysis conducted together with Loet Leydesdorff and reported in a forthcoming article, I found that Jaccard’s measure did not serve this analysis as well as the Salton Index. Jaccard’s Index provides the intersection of the two countries as a percentage of the sum, while the Salton Index provides the intersection as a weighted percentage. But the difference is more than a factor two: whereas the Jaccard Index focuses on strong links in segments of the database (e.g., the strong relations between Croatia and Slovenia), the Salton Index organizes the relations geometrically so that they can be visualized as structural patterns of relations (Hamers *et al.*, 1989 explores this in detail). Unlike the Pearson correlation, however, the Salton Index remains non-parametrical (Ahlgren *et al.*, 2003; Leydesdorff & Zaal, 1988). Thus, in keeping with Glänzel (2001), the Salton Index is used as a measure of the networked relatedness of countries.

Following analysis of the data using the cosine, a factor analysis was conducted on the data. Factor analysis is used in two ways in this project: one to factor analyse the global data sets and two, to analyse the citation relationships among journals for the case studies. Factor analysis is a mathematical tool that can be used to examine a wide range of data sets. It allows the analyst to detect structure in the relationships between variables at the multilateral level. Together with Loet Leydesdorff, I applied factor analysis to the asymmetrical matrix of co-authorships created using the SCI addresses from the 2000 and 1990 CD-Rom data. The data was factor analyzed forcing different numbers of factors to reveal different information about the structure and architecture of the relationships. In
comparison to the observed data discussed above, factor analysis enabled us to reveal structural properties of correlation and variation that are not observable by inspection of the matrix level. This was the case with both the global network and with the citing relationships of journals, described below.

4.4 Case Studies of Global Dynamics in Disciplines of Science

The global network of science presents intriguing information on the network of scientists linking with each other at geographic distances. In order to explore the nature of this network further, six case studies were conducted at the sub-field level. In order to explore the network structures of disciplinary fields, I conducted a series of cascading analyses, illustrated in Figure 4 as a series of steps, processes, and products. The first used the citing relationship among existing journals to uncover the cluster of journals that scientists identify as part of their field. Secondly, for several years within the decade of 1990 to 2000, I collected all the addresses of the articles in the relevant journal clusters. I then used this in the third step to analyse the percentages of international co-authorship within these years to see the extent to which the field had grown as an international discipline. Finally, in the fourth step, I created networks of linkages among scientists at the national level. In the final step, I analyzed the networks to test whether the mechanism of preferential attachment could be shown to be operating within the network of authors at the sub-field level and to test for “small world” properties of these networks. Each step is described in more detail below.
The first part of the analysis involved identifying fields of science for study. The first question here was how to chose the fields. There have been several attempts at classifying meta-fields of science, including placing scientific inquiry into classes such as “observation,” “experimental” and “theoretical.” (Zuckerman & Merton, 1971) However, for the purposes of this analysis, these existing categories did not allow me to group areas of science adequately to ask questions about communications at the global level. Accordingly, I chose to characterize the sciences based upon access to resources instead of solely by method.

Acknowledging that there is overlap, and that categorization requires some judgment on my part, for the purposes of this study, scientific inquiry is categorized principally into one of three categories of research making up a typology of the sciences. These are 1) data driven research, 2) resource driven research, 3) equipment driven, and 4) theory driven research. These categories help identify the factors affecting the organisation of research, as well as ways in which knowledge is created. These categories are useful for discussing international cooperation because the factors are often cited as reasons for seeking collaborative ventures.
The table below provides a notional concept of this typology, showing different fields of science considered for this part of the study, as they would be placed within this categorisation.

<table>
<thead>
<tr>
<th>Data-driven collaboration</th>
<th>Resource-driven collaboration</th>
<th>Equipment driven collaboration</th>
<th>Theory-driven collaboration</th>
</tr>
</thead>
<tbody>
<tr>
<td>biomedical</td>
<td>oceanography</td>
<td>high energy physics</td>
<td>mathematics</td>
</tr>
<tr>
<td>genetics</td>
<td>geology</td>
<td>astronomy</td>
<td>economics</td>
</tr>
<tr>
<td>demography</td>
<td>seismology</td>
<td>energy</td>
<td>sociology</td>
</tr>
<tr>
<td>computer</td>
<td>zoology</td>
<td>avionics</td>
<td>polymers</td>
</tr>
<tr>
<td>epidemiology</td>
<td>soil science</td>
<td>space</td>
<td>science studies</td>
</tr>
<tr>
<td>virology</td>
<td>anthropology</td>
<td>manufacturing</td>
<td>philosophy</td>
</tr>
</tbody>
</table>

Table 6. Lists of fields of science by characteristics driving collaboration

Within these characterisations are implicit assumptions, ones with theoretical implications: this is discussed in Chapter IV. By indicating the drivers for collaborative ventures, I am also imposing a testable hypothesis on the organisation suggested for this study, one that will be discussed in the chapters describing the case studies (Chapters III, IV, and VII).

4.5 Defining the journals for analysis in the case studies

Research shows that scientific journal-journal citations can be exploited to expose patterns of interrelationships within and among fields of science. Cozzens & Leydesdorff (1993) suggest that disciplines of science can be operationalized in terms of journal sets. As such, journal-journal citations can be used to track changes in the disciplinary structure of science. The patterns provide a method for studying the structure of a field in a single year as well as over time. The concepts and tools developed by Leydesdorff and Cozzens were applied in Step 1 and its accompanying process illustrated in Figure 4.

The citation relationship among journals reveals a structure of the literature that scientists view as relevant to their work. An analysis of these citation relationships reveals "clusters" that can be visually depicted. This method of journal-journal citation mapping is applied in this study to see if the structure of citations related to the fields has changed, how they are related to other fields, and whether international linkages have influenced the field.

Using the method developed by Leydesdorff & Cozzens (1993) and later applications by Leydesdorff, Cozzens, and Van den Besselaar (1994), I identify the
journals most closely associated with the field. In describing the methodology, Leydesdorff and co-authors note out that journal-journal citation relations “contain information about field and subfield structures at a sufficient level of aggregation for the construction of indicators…” (Leydesdorff et al. 1994) Journal publications are not stable over time: The incremental change can become an indicator in itself by revealing the structure of the field over time. For example, the inclusion of new journals within a cluster of related journals may indicate the extension or further differentiation of the clusters or it may indicate qualitatively new developments. The emergence of new clusters, or the merging of clusters, can also be indicative of structural changes in a field.

The Leydesdorff-Cozzens method enables the creation of journal-journal citation maps based on factor or eigenvector analysis of the citations of journals in the Science Citation Index (SCI) and the Journal Citation Reports. The data I report here are drawn from the databases of the Institute for Scientific Information SCI and JCR CD-ROM data. The data consist of listings of all journals processed and included on the CD-Rom for 1994 and 2000, with a comparison of the shape of the field in 1990 using the Web of Science online database. Co-author names and affiliations are obtained from the Science Citation Index extended available through the Web of Science in between April 2002 and April 2003.

4.6 Revealing the structure of the field

The first step of identifying initial journals to work with involved using the Web of Science to identify journals with titles related to the field in question. Leydesdorff & Cozzens (1993) have proposed “central tendency journals” as yardsticks for measuring structural change in fields of science over time. Central tendency journals are defined as seed journals that exhibit the highest correlation with the eigenvector that represents the cluster at the network level. Central tendency journals exhibit more stability than journals that are less central to the cluster.

Thus, for example, in the case on seismology, I searched for the words “earthquake” and “seismology” in journals available in March 2002 on the Web of Science to identify journals closely related to this subject. A number of journals were initially considered, although two journals had names that appeared directly relevant to the case: The Journal of Seismology and The Bulletin of the Seismological Society of America. Using a factor, it can be observed that that, in 2000, the Bulletin of the Seismological
Society of America (BSSA) is the journal most closely associated with the field of seismology. BSSA thus served as the central tendency journal (CTJ) for the case.

The citing pattern of the journal set was further analyzed to identify the action parameter associated with the field. Journals being cited by articles in CTJ indicate the structure of the field as viewed by the authors. For the purposes of the case studies, the set of citing journals for each year is used. The central tendency analysis is used to show local densities in the overall network for each year. Clusters can be compared over years, even though they are not necessarily tied to the same CTJ. This helps illustrate the evolution of the field over time. The evolutionary illustrations show the relations within the cluster over the years by showing the relevant citation environment. The change in the data set can be used as an indicator of restructuring within the field.

The clusters of journals that emerge from the citing patterns around the CTJ in the year 2000 reveal the structure of the field in that year. To provide an example, Table 7 displays the results revealing the BSSA as the "central tendency journal" for 2000 as well as its citing galaxy, highlighted with a box. Seismology loads on factor 2.
Table 7: A factor analysis of the citing patterns of the Bulletin of the Seismological Society of America, 2000

---- FACTOR ANALYSIS ----

Varimax Rotation 1, Extraction 1, Analysis 1 - Kaiser Normalization.

Varimax converged in 5 iterations.

Rotated Factor Matrix:

<table>
<thead>
<tr>
<th>Variable label</th>
<th>Factor 1</th>
<th>Factor 2</th>
<th>Factor 3</th>
<th>Factor 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>V1635</td>
<td>.96755</td>
<td>.13346</td>
<td>.03939</td>
<td>.00451</td>
</tr>
<tr>
<td>V3197</td>
<td>.92073</td>
<td>.09846</td>
<td>.11211</td>
<td>-.04233</td>
</tr>
<tr>
<td>V2112</td>
<td>.89984</td>
<td>.08973</td>
<td>.19592</td>
<td>-.05187</td>
</tr>
<tr>
<td>V2110</td>
<td>.86183</td>
<td>.25644</td>
<td>.00527</td>
<td>.16200</td>
</tr>
<tr>
<td>V4661</td>
<td>.82490</td>
<td>.05293</td>
<td>.08132</td>
<td>.08248</td>
</tr>
<tr>
<td>V4897</td>
<td>.82072</td>
<td>.47634</td>
<td>.03008</td>
<td>.06451</td>
</tr>
<tr>
<td>V5392</td>
<td>.74253</td>
<td>.15776</td>
<td>.04004</td>
<td>.09377</td>
</tr>
<tr>
<td>V734</td>
<td>.59009</td>
<td>.60293</td>
<td>-.19081</td>
<td>.19118</td>
</tr>
<tr>
<td>V5231</td>
<td>.18917</td>
<td>.89607</td>
<td>-.07031</td>
<td>-.23901</td>
</tr>
<tr>
<td>V3637</td>
<td>.38826</td>
<td>.84096</td>
<td>-.03825</td>
<td>-.09281</td>
</tr>
<tr>
<td>V4217</td>
<td>.60390</td>
<td>.71365</td>
<td>-.01758</td>
<td>-.02778</td>
</tr>
<tr>
<td>V398</td>
<td>.11106</td>
<td>.53651</td>
<td>.01291</td>
<td>.06764</td>
</tr>
<tr>
<td>V1728</td>
<td>.08075</td>
<td>-.06509</td>
<td>.93182</td>
<td>.00392</td>
</tr>
<tr>
<td>V5139</td>
<td>.03365</td>
<td>-.08399</td>
<td>.93085</td>
<td>.00752</td>
</tr>
<tr>
<td>V4231</td>
<td>.12658</td>
<td>.08081</td>
<td>.86149</td>
<td>-.02022</td>
</tr>
<tr>
<td>V1500</td>
<td>-.01183</td>
<td>.06809</td>
<td>-.15294</td>
<td>.80088</td>
</tr>
<tr>
<td>V2113</td>
<td>-.16112</td>
<td>.24468</td>
<td>-.18170</td>
<td>-.62480</td>
</tr>
<tr>
<td>V1635</td>
<td>.96755</td>
<td>.13346</td>
<td>.03939</td>
<td>.00451</td>
</tr>
</tbody>
</table>

Variable labels:

V398 'ANN GEOLIS
V734 'B SEISMOLOG SOC AM
V1500 'CURR SCI INDIA
V1635 'EARTH PLANETS SPACE
The factor analysis for the citing relationships results in a plot of the stimulus space showing a 2-dimensional drawing of the citation relationships. The next section displays and discusses the figures that emerge from this analysis.

4.7 Mapping the international network of the field

Each of the journal sets identified were analysed by calculating the percentages of international co-authorship as a share of all papers published in that year for that cluster. Then, I mapped the network of the field in both 1990 and 2000. This is done using the following techniques:

1. For each of the cluster of journals, I collected all the related records from the Web of Science for that year. I saved these records with authors and addresses so that I could identify linkages among authors by country based on the addresses. (Step 2 in Figure 4.3.)

2. Using custom computer programs written for this purpose, I sorted the author names and addresses into files that allowed me to count international co-authorships and figure the percentage of all papers that are internationally co-authored. (Step 3 in Figure 2.)

3. Taking the address file created as part of this analysis, I imported this into UciNet to create an affiliations file, and then this data was exported to Pajek network analysis software to draw a social network. The network was analysed and displayed within Pajek using the Kamada-Kawai equation. A core analysis was conducted on
this network to reveal the most intense relationships among countries. The sets of resulting graphs are saved first in bitmap format and then in Word; they are presented in Appendix A.

4. Importing the data set into SPSS, I used cosine analysis to explore the hierarchy of the network, revealed by using Pajek network analysis software. Within the cosine analysis, in Pajek, I also conducted a core analysis showing the most intense relationships within the core group. The sets of resulting graphs are saved first in bitmap format and then in Word; they are presented in Appendix A.

5. Then, I conducted a power law analysis and a small world analysis on the networks of authors and co-authorships to see if the networks contain properties that would suggest they are part of a self-organizing network.

Once the structure of the field is revealed and understood for 2000, the next question is how the field has changed over time. Using the structure of the field in 2000, and retaining the CTJ identified in 2000, I reconstructed the field historically for 1990 and 1994. The concept of extending back historically rather than beginning in 1990 and working forward is so that I can examine the evolution of the field over time. If one systematically accounts for delineation in the groupings ahead of time, one risks making a prediction of performance with reference to an outdated unit. The point is to identify the dynamics of the field: The validity of this method is demonstrated graphically from the seismology case by the fact that the journal Seismology, which would appear to be the most relevant journal, was only initiated in 1999. If I had fixed the journal set only to those available in 1990, I would have missed the entrance of the new journal. So I began with the structure of the field as it is revealed in 2000 and reconstructed the development from the perspective of citing relationships in that year.

4.8 Determining the network structure of the fields of science
Once the networks were revealed at the global level and within fields at the international level, it became clear that to really understand the dynamics of knowledge creation, it would be important to understand the structure of the relationships among the researchers themselves. To explore this, I did two things: one is to examine the author networks within the case studies, and secondly, I conducted a series of semi-formal in-depth interviews. This section describes the analysis of the author relationships. The next section describes the methodology for conducting interviews.
As part of the case study analysis, authorship data for all of the 19,147 articles within 65 journals have been collected for the SCI 2000 publication year. All papers within the journals have been included in the data set: no limitations regarding document type, language or any other attributes were applied. Author names were taken as recorded into the database; no attempt was made to adjust for spelling variants. This data was then analyzed for author frequency within each case, and the resulting curves suggested that further analysis of the possibility of small world properties and scale-free power laws within the networks would be worthwhile.

To conduct this analysis, the data set created by the author names was further analysed to identify those who had co-authored papers at the international level. This subset of authors became a new dataset, and the occurrence of international co-authorship was placed into a pivot table to count the number of times that an author had co-authored papers with other colleagues in the data set. The resulting occurrence table was then exported into Ucinet for several sets of analyses: 1) degree centrality and 2) cluster coefficients. This was then made into a graph, converted to a log-log distribution, and analysed for the mechanism of preferential attachment. The results of this analysis are presented in Chapter VII.

4.8 Interviewing Scientists: Understanding Motivations and Working Relationships

Bibliometric and network analysis provides a significant, quantitative picture of the patterns of international collaboration. In order to gain a richer and more nuanced picture of international collaborations, as well as to gather a sense of the motivations for and methods of collaborative work, I conducted a series of interviews with scientists involved in international collaboration. The method of conducting the survey involved a series of telephone interviews.

The subjects to be interviewed were identified using a series of analytic tools described above, where I identified the frequency with which a researcher co-authored with someone in the data set:

1. First, the list of all the authors publishing in the specific field of interest (see method described above) was derived from the full database of all articles (year 2000) gathered for the case study from the Web of Science. A program designed to separate the names of authors from the full file was used and a new Excel file was created.
2. Once all the authors were in a single database, I ran a special customized program to identify those authors who had internationally co-authored an article.

3. With this list, placed in a spreadsheet, I created a pivot table in Excel to produce a frequency analysis of how many publications could be attributed to a single author.

4. Using this data, I conducted a frequency analysis in SPSS and sorted the data by the author with the highest number of publications at the international level, based on publications. (This data was charted and analysed separately, and the results will be described below.)

5. For the top 20 authors in each field, I identified an email address for that person, and I sent them an invitation to participate in a telephone interview. For an additional 30 key scientists, I also mailed them a letter and a description of the project to invite them personally to participate.

6. For those scientists who agreed to be interviewed, I undertook a phone conversation with them based upon two things:
   a. I selected an article by this scientist that was published in the year 2000 where there were international co-authors, and I used this article on which to base the discussion. If there was more than one article by that author, I chose the one that had a higher citation rate or that included authors from more than 2 countries.
   b. A set of questions guiding the discussion, included below.

The interviews sought to elicit from scientists their understanding of the reasons an international collaboration took place, how information was exchanged, how new knowledge was created and shared, where the research took place, how the results of research were disseminated, and whether follow-on work took place. Scientists were asked to assess whether they think of the collaboration as “successful,” and what scientific benefit derived from this collaboration.

The following questions guided the discussions:

1. As a way to begin the discussion, let’s focus on one of the papers you co-published in 2000. How did this partnership develop? Are you long-time colleagues or did you come together for this project only?

2. How long did the partnership exist before you published an article together?
3. Would you say that you and your collaborator work in the same field of science? (If no, what field does your collaborator work in?) Do you have complementary capabilities?

4. What was the initial communication tool you used when you were first conceptualising this project? (mail, phone, internet, face to-face meeting?)

5. How did you maintain communications during the course of the project? What methods did you use to share data?

6. Did you choose this collaboration over other opportunities to conduct research? If so, how did you decide on this one?

7. How was the research funded? Did both sides bring funding to the project?

8. If the researcher partners each funded their own work, were the funding levels equivalent in each institution? (6.a. Even if funding differed, sometimes in-kind contributions can make up the difference. Did your partners offer in-kind contributions to the research?)

9. What was the nature of the research? (e.g., observation, experimentation, theory, development, evaluation, testing)
   a. At what location(s) did field work (or experimentation) take place?
   b. At what location(s) did analysis (or development) take place?
   c. Did the remote locations of part of this research affect your productivity?

10. Did the collaborative nature of this research affect the way you conducted your work? If so, how?

11. What were some of the motivating factors for maintaining momentum in the research?

12. How did you share the task of writing the findings of your research for publication?

13. Are you planning to work with this partner(s) again?

14. Overall, do you think that the international collaboration was successful? Why?

In total, 25 in-depth (more than 30 minutes) telephone interviews were conducted. Although I had planned to conduct a broader survey, I found it difficult to obtain interviews with scientists in certain fields. For example, while I was conducting this research, I attempted to contact researchers in the field of virology. Due to the outbreak of Severe Acute
Respiratory Syndrome (SARS), international researchers were very difficult to contact. (I ultimately spoke to only 3 virologists for this research.) In addition, the number of researchers working at the international level in polymer research and seismology turned out to be quite small compared to other fields, so a smaller number of people were available to be interviewed than I first expected.

4.9 Determining the capacity of countries to participate at the global level

It is possible that the supply of researchers from diverse countries available to collaborate is one of the reasons there is more international collaboration. To provide a baseline for understanding this, together with two RAND colleagues, I updated an index of science and technology capacity that we had created for the World Bank a few years ago. This capacity index is included as Chapter VI. To create the index, data was collected from eight different data sources for 115 countries of the world. Complete data was available for 76 countries. For these 76 countries, the data was indexed and compared, creating a comparative assessment of how countries measured up to an international mean. A parsing scheme is suggested in the article for categorizing countries based on their revealed capacity. This process is described in more detail in Chapter VI.

5. Organisation of the Thesis

This introduction provides the broad overview of the research questions, definitions, and methodologies used in the study. Some of the material is repeated in the chapters, and I apologize for the repetition, but it seemed important to provide a broad overview of the methodology in the introduction. Each chapter except this one was prepared as a stand-alone article for submission to a journal. As a result, each has a slightly different format. Each chapter that was prepared for journal submission also has a bibliography, and then there is a full project bibliography at the end of the thesis that also repeats the bibliographic entries in the chapters. This is done to ensure completeness, since some references will only appear in the final bibliography.

The thesis is organized to bring the reader along a transition from understanding first the policy questions and issues surrounding international collaboration in science. This helps to set the stage for the discussion that unfolds in the succeeding chapters. Chapter II, “Science and foreign policy: the elusive partnership,” sets out the questions and issues surrounding international science to highlight the need for much greater understanding of the dynamics of international collaboration.
The exploration of the dynamics begins with a discussion of the network of collaborations at the field level by first examining collaboration within seismology. Chapter III, "Seismology as a case study of dynamic, distributed science" presents a bibliometric analysis of seismology in 1990 and 2000. This is an interesting case study because seismology is shown to emerge as a separate field of science from the larger field of geophysics. It did not exist as a separate field in 1990. This gave me a chance to test whether newer, emerging and specialized fields of science are more highly internationalised than other fields.

The next chapter in the thesis presents the results of six case studies of international collaboration in science. The six cases presented in Chapter IV are: astrophysics, geophysics, mathematical logic, polymer science, soil science, and virology. Collaborative networks within these fields are examined in 1990 and in 2000 to identify the extent of growth in international collaboration in these fields over the ten-year period. Part of the goal of this research was to test whether the drivers behind international collaboration were different in different fields, and whether these drivers affected the growth in international collaboration.

Chapter V provides an overview of networks of collaboration at the global level. This helps to provide the reader with a high-level view of the structure of international collaboration and how these patterns changed from 1990 to 2000. The data are examined from a number of perspectives, including observed and normalized networks, as well as a factor analysis of the hierarchy of collaboration. An appendix will provide all the networks that were created for the project at the observed and normalized levels.

In an effort to understand the conditions that can influence the growth of global dynamics, Chapter VI presents an index of science and technology capacity for 76 countries of the world. This index gives insight into the extent to which countries have developed enough scientific research capacity to participate in world-class or international research. It discusses how capacity contributes to knowledge creation at the local and the global levels.

In order to bring many of the different threads of research together, Chapter VII presents a comprehensive approach to international collaboration in science by suggesting a theory to explain the rapid increase in international collaboration in science. Bringing together the pieces of analysis from the global level, from the case studies, from the analysis of author frequencies and linkages, this chapter marries network theory and the data presented in the earlier chapters to offer a theory of the growth of international collaboration that fits with the data created for this project.

34
The conclusion presented in Chapter VIII discusses the extent to which the project has supported or refuted the expectations set out at the beginning of the research. It also discusses the unexpected findings as well as the limitations of this research. Policy implications of this research are briefly discussed.

Appendix 1 presents the observed networks created for the six case studies of international collaboration discussed in Chapter IV. Appendix 2 presents the data used to create the science and technology capacity index in Chapter VI.