Exploring scientific collaborations in Geographical Information Science (GIScience): A study of its co-authorship networks

Maria Cristina Arciniegas-Lopez

Thesis submitted in fulfilment of the requirements for

PhD in Information Science

City University, London
Department of Information Science
Schools of Informatics

July 2007

Declaration on consultation and copying

I grant powers of discretion to the University Librarian to allow this thesis to be copied in whole or in part without further reference to me. This permission covers only single copies made for study purposes, subject to normal conditions of acknowledgement.

Maria Cristina Arciniegas-Lopez, May 2007

Table of Contents

List of Figures

List of Tables

List of Equations

Acknowledgements

I wish to acknowledge my particular indebtedness to my supervisor Jo Wood for his continuing guidance, constructive criticism, inspirational discussions and supervision during all stages of this work. I would like also to thank all the people at giCentre - City University, especially Peter Fisher, Jason Dykes, Jonathan Raper, Vesna Brujic-Okretic, David Mountain, Anil Gunesh and Fotis Liarokapis for their advices, and constructive suggestions. To Andrej Mrvar, Pajek software developer, for his invaluable help and advices on how using his software tool.

To my friends at City University Steph Marsh, Panayiota Polydoratou, Jutta Haider, Waraporn Jirapantong, Gilberto Cysneiros, Vlasios Voudouris, and Hamid Jamali for their unconditional friendship and our endless discussions about PhD issues.

I should also like to thank my mother Esther Lopez, my brothers Fausto and Gustavo Arciniegas and all my family in Colombia for their affection and moral support . To my friends Mila Gamboa, Marta Cabrera and Olga Castillo for their unconditional support and understanding. Finally, I must to thank and express my gratitude to my husband Pedro Hernandez for his love, endless patient, unconditional support and encouragement throughout the PhD time. I would not be able to reach this stage without him.

Abstract

Geographic Information Science (GIScience) as a discipline focuses on fundamental issues surrounding Geographic Information (GI) and developments and applications of Geographical Information Technologies (GITechnologies). GIScience has accumulated a body of knowledge that can be easily exported and applied to other disciplines and assembled a wider multidisciplinary research community.

Co-authorship networks are used to explore GIScience scientific collaborations during 1992-2002. Six different co-authorship networks were built from publication outlets comprising different sets of core and peripheral journals. The closer the periphery to the core, the more relevant the selected journals are to GIScience. Topological characteristics of all networks show similar networks despite the differences in sizes and the nature of the topics covered. However, networks with the peripheral journals closer to the core were more centralized around well-known scholars within the discipline. Furthermore, the network structures show a GIScience core linked to allied disciplines, especially to a highly clustered remote sensing research community.

The core co-authorship network was geo-referenced using authors' affiliation information. The results show that geographical proximity, language and cultural preferences play important roles. Countries known for their strong publishing patterns in other sciences such as England, USA and Canada were alos identified within GIScience domain. A growth of international collaboration among Scandinavian and European Countries was revealed. Results also show that China, India and Brazil have been increasing their international participation within the GIScience research community.

1 Introduction

Geographic Information (GI) is everywhere and in many formats. From traditional road paper maps, high risk or disaster areas defined using satellite imagery, to real time position information obtained from global positioning system (GPS). However, it was during the sixties that the first computational tools (Geographical Information Systems) to handle digital GI were developed. Back then GI posed challenges regarding storage and manipulation, mainly limited by the computational and hardware restrictions of the time. Nowadays, as digital GI has become more accessible and is exported to many other disciplines, new types of geographical information technologies (GITechnologies) have emerged. Therefore, new sets of challenges spanning many disciplines have arisen, and the multidisciplinary research community addressing them has become more noticeable. In this context, Geographic Information Science (GIScience) has emerged as the branch of Information Science that deals with the fundamental set of principles underlying the design, testing, and use of GITechnologies (Goodchild, 2004). However, the field's disciplinary boundaries have become fuzzy and the research community more diffuse as the relevance and the applicability of GI and GITechnologies span more disciplines, and fundamental issues and advances in other disciplines such as geography, computer science, statistics or cognitive sciences become more relevant to GIScience. This thesis is an attempt to delineate GIScience through its network of scientific collaboration.

The overall aim of this thesis is to build collaboration networks that better represent and allow exploring the multidisciplinary nature and fuzzy disciplinary boundaries of GIScience.

1.1 What is special about GIScience?

Forty years ago, the emergence and development of GISystems was the answer to the lack of capability of traditional information systems to handle GI and associated spatial keys (Goodchild, 1992). The ability of GI to represent continuous surfaces using unlimited numbers of <x,y,z> tuples that generate very large files required the use of special spatial-oriented analysis tools (Longley *et al*, 2001). Besides, the characterization of complex geographic landscapes required the development of new conceptual data models and special tools that incorporate spatial dependence and generalization methods (Goodchild *et al* 1999). Moreover, the GIScience and allied research communities face unresolved challenges such as developing frameworks for handling time and the third dimension properly (Goodchild, 1992; Longley *et al*, 2001), overcoming the limited representation of a geographical phenomenon on a flat surface (Wright *et al*, 1997) and a deeper understanding of societal issues surrounding GITechnologies and GISystems usability (Goodchild *et al*, 1994). All these issues are of enough significance and complexity that they cannot easily be placed under any existing disciplines other than GIScience (Goodchild, 1992). In addressing these challenges, the GIScience research agenda focused on three distinct areas that involve individuals as users of the technologies; the systems represented by the technology per se (hardware, software and distribution tools); and the society and its institutions that implement, use and need the technologies (Goodchild, 1999).

GIScience needs to solve complex questions that mainly require the knowledge and principles of other sciences, but also other disciplines have embraced GIScience's concepts and methods to address their own spatial-related problems. Therefore, when wishing to communicate their research results and findings, scholars working on GIScience issues may chose publication outlets not only relevant to GIScience but to other disciplines. Consequently, representing GIScience through its collaboration networks is a challenging task because its particular characteristics have structured a discipline with focus on both fundamental and applied research, yielding multiple and diverse publication outlets and a research community with multidisciplinary backgrounds and carrying out research on cross-disciplinary topics.

Three research questions delineate the scope of this study.

1. How to delineate GIScience in terms of journals?

Generally, traditional disciplines such as mathematics, physics or chemistry possess a well-defined list of journals listed under distinguishable subject categories on bibliographic databases (Newman, 2004b) that cover the most important findings in the subject. However, there is no comprehensive list of journals of interest that can be used to represent GI Science (Fisher, 2006).

Journals such as the *International Journal of Geographic Information Science (IJGIS)* and *Transactions in GIS* are identified as umbrellas where original ideas, techniques and approaches regarding applied and fundamental research findings can be found (Fisher, 2006). Therefore, they can be considered the core journals at the centre of the discipline publication forums. In other to find out the discipline's other relevant publication outlets, one could take the other journals classified in the same subject category as the core journals.

For instance, IJGIS is placed in different categories depending upon which bibliographic database is examined. Ingenta-Connect classifies IJGIS under *Earth and Environmental Sciences* with 397 other journals, EBSCOHost classifies IJGIS under *Geography, Anthropology and Recreation* with other 219 journals and finally, the Social Sciences Citation Index (SSCI) places IJGIS under two categories: *Geography* with 40 others journals and *Information and Library Science* with 59 journals more. All these classifications group journals that represent a very wide subject area. Thus, one cannot assume that research collaboration in GIScience can be accurately represented by all journals under the same categories as GIS. Hence, traditional subject classifications used by bibliographic databases are not convenient to represent GIScience.

This study attempts to define the publication outlets that better define GI Science. In doing so, it proposes identifying a 'core', and then, selecting other peripheral journals by quantifying the distance between core and periphery in relation too the amount of work that cross-disciplinary authors have published. Core journals cover a wide variety of related topics from contributors working on a wide set of research areas, many of them, as Arciniegas and Wood (2006) argue, may not choose to place themselves as GIScience scholars. On the other hand, peripheral journals to IJGIS may be core or specialized forums from allied disciplines or related areas. These journals may be visible to specific scientific communities that are linked to the core by authors with cross-disciplinary research interests. By representing collaborations in GIScience by co-authorships in core and peripheral journals, one avoids drawing boundaries around GIScience and its allied disciplines.

2. How should scientific collaborations in GIScience be characterized?

This study attempts to characterize scientific collaborations in GIScience by exploring the structure of co-authorship networks made up by linking co-authors of scientific papers from the discipline's publication outlet. As the structure of a co-authorship network can be used to depict the academic society behind a discipline (Newman, 2001b), by investigating its structure one can reveal the impact that individuals and their scientific contributions have had on the discipline (Arciniegas and Wood, 2006). By representing research collaborations as a network the result of social-academic interactions between scholars publishing papers can be measured by concepts borrowed from three fields:

- Co-authorship measures from Bibliometrics to quantify research outputs and patterns of collaboration and co-authorship.
- The notion of distance from Graph Theory to identify central authors and any especial topological characteristics of the co-authorship network.
- Centrality measures from Social Network Analysis to measure the impact of the authors' participation and publication patterns in the network structure.

Additionally, by exploring the geography of the GIScience research collaborations, this study attempts to analyse the geographical distribution of the collaboration network.

3. How should multidisciplinarity be measured?

Multidisciplinarity is one of the agreed characteristics of both scholars and research around GIScience (see question number one). One expects that the publication outlet comprising core and peripheral journals covers a wide variety of research topics from allied and very close disciplines to GIScience. However, each field (represented by the journals' research topics) will contribute differently in shaping the structure of the co-authorship network. This study proposes using statistical entropy (SE) as a measure of the variation or diversity (Börner *et al*, 2005) of the authors' publication sources in the discipline's publication outlets (see chapter 5). Thus, through the analysis of the distribution of authors' SE, it would be possible to identify to what extent the network is built from collaborations concentrated on few journals, which would show levels of a highly specialized research community. Alternatively, if built from more evenly distributed collaborations around a higher number of journals in the publication outlet, it would describe the multidisciplinary nature of the discipline. .

The overall aim of this thesis is therefore to build collaboration networks in GIScience that better represents the multidisciplinary nature and fuzzy disciplinary borders of the field and allow exploration of its characteristics. In order to achieve this aim, this study pursues the following objectives:

- To assemble the publication outlets that take into account the distinct characteristics of research in GIScience.

- To build a co-authorship network that can be taken as a window on the structure of GIScience by considering co-authorship as a surrogate measure of research collaboration.

- To evaluate the nature of the GIScience research collaboration network measuring and analyzing basic properties such as co-authorship, collaboration and authors' participation distributions, and topological features such as centrality measures and network distances and connectivity.

- To evaluate the role of geographical proximity in the formation of GIScience collaboration networks.

1.2 Thesis Outline

Chapter 2 reviews and discusses the theoretical context surrounding this study. The first part discusses the latest attempts to evaluate scientific activities using bibliometric measures, especially co-authorship and its main limitations. The second part explores the literature surrounding the representation of networks of any type using the graph paradigm. The third part explores the relevant literature covering the evolution of the network structures, from ordered lattices to random structure, small worlds and ending with scale free networks. Finally, the fourth part presents the latest approach to understand large scale networks deploying concepts of social networks and graph theory.

Chapter 3 presents the approach proposed for this study to describe GIScience collaboration networks, tailored to the field's highly multidisciplinary nature. The first part of the chapter discusses GIScience as a discipline, interpreting its multiple definitions as a field, exploring its unique characteristics and its model of knowledge production. The second part presents the proposed delineation of GIScience in terms of academic journals used to document its development. The last part explains in detail how the co-authorship networks that represent GIScience were assembled, including the computational tool developed to manage more than 60,000 bibliographical records, and the problems encountered regarding the lack of author-name standardization.

Chapter 4 is devoted to exploring the structure of six multidisciplinary co-authorship networks. The emphasis is placed on measurement patterns of network participation, co-authorship and collaboration that allow characterization of research in GIScience and also to compare the field with other disciplines. The second part of the chapter identifies the network core, evaluating local authors' interactions through various centrality measures. The third part explores the use of component sensitivity analysis for establishing to what extent the network structures are centralized.

Chapter 5 describes the multidisciplinary nature of the GIScience co-authorship network. The emphasis is placed on understanding multidisciplinarity through the patterns of collaboration, participation and co-authorship from works in the publication outlets. Statistical Entropy is used to measure the degree to which an author's work is concentrated or distributed across the journals.

Chapter 6 explores the geography of the GIScience co-authorship network. The emphasis here is to develop a method to georeference the scientific work using authors' national affiliation addresses. The second part explains in detail the computerized tool developed to make the country-collaboration pairs and also presents the observed ISI-WoK limitations that affected the georeferencing process and the proposed solution. Finally, Salton measures are used to establish the strength of the collaboration ties between countries, revealing the geography of GIScience research.

Finally, chapter 7 draws the conclusions of this study and re-assesses its aims and objectives. It also presents the advantages and discusses the disadvantages, giving guidelines for further work.

A computational tool was developed to aid the process of building and geo-referencing the co-authorship network, and also to implement the proposed solution to the author names standardization problem. However, producing a computerized tool was not an objective of this study, thus, only the most relevant lines of code are included in the thesis.

2 Research Context

The aim of this chapter is to review and discuss the theoretical context that surrounds this thesis. The chapter exposes the multidisciplinary nature of the knowledge needed to achieve the thesis objective as analyzing the structure of a scientific collaboration network, especially a burgeoning and highly dynamic one such as GIScience, which embraces methodologies and concepts from fields as diverse and dissimilar as mathematics, sociology, information and computer science and physics.

2.1 Evaluation of Scientific Activities

Scientific communication is a highly complex problem that according to Mikhailov (1984) can be defined as the combined process of presentation, delivery, and receipt of scientific information in human society. As part of the scientific communication process, scientific or research collaborations are social processes governed by the complexity of human interactions (Bordons and Gomez, 2000). Thus, the precise definition of scientific collaboration, the analysis and description of its different types or the quantification of the contribution of each participant are very challenging tasks (Bordons and Subramanyam, 1983; Katz and Martin, 1997; Gomez, 2000).

Nevertheless, it has long been realised that published scientific works provide a potential window on the collaboration of scientists (Garfield, 1955; Price, 1965; Melin Persson, 1996; Bordons and Gomez, 2000; Newman *et al*, 2006). To take advantage of the coded data about authors and their works within the literature, scientists developed a family of techniques referred as bibliometrics to identify relationships within the published literature. In this way, bibliometrics approaches such as citation analysis, content analysis or co-authorship analysis have been common tools for information scientists to study the structure and process of scientific communication and collaboration (Subramanyam, 1983; Melin and Persson, 1996; Ding *et al*, 1999; Borgman and Furner, 2002).

2.1.1 Bibliometric Measures

The term bibliometrics was defined by Pritchard (1969) as "*the application of mathematics and statistical methods to books and other media of communication*". Bibliometrics, instead of focusing on the physical properties of documents, draws the attention to statistical patterns in variables such as authorship, sources, subjects, geographical origins, and citations. Von Unbegrn-Sternberg (2000) emphasises the fact that bibliometrics is particularly related to research in scientific communication, and the publishing system that is one of the bases for studying this communication process.

In that way, bibliometrics has been developed through the use of quantitative tools for study physical published units or bibliographic units (Broadus, 1987). Despite the fact that quantitative analysis of scientific publishing began in the 20s and 30s, when the mathematical models behind Lotka's law of scientific productivity, Bradford's law of scatter journals and Zipf's law of words were developed, bibliometric methods were not used for measuring scientific activities until the 60s, when content-based and technical tools needed for bibliometrics where developed (Bordons and Gomez, 2000).

After the 60s, authors agree that bibliometric data was particularly useful for studying longitudinal trends in scientific disciplines as larger data sets became available and analysed. Among the advantages of bibliometric methods is the fact that it is an unobtrusive data collection process (Subramanyam, 1983) from which any results can be easily replicated by others using the same procedures and measures. Moreover, bibliometric methods have been deployed and recommended by many outside the information science research community such as mathematicians, physics and computer scientists (Newman, 2001a; Barabasi *et al*, 2004; Börner *et al*, 2004; Moody, 2004; Murray *et al*, 2004; Newman, 2004c).

Therefore, as widely used tools, bibliometric methods have supported the study of the development of science over time, its content, its geographical and organisational distribution and the effects of co-operation on research (Borgman and Furner, 2002; Murray *et al*, 2004; Leydesdorff, 2006a). However, Von Unbegrn-Sternberg (2002) stresses the importance of interpreting bibliometric results within the context of each specific study. Although it is clear that bibliometric indicators can give signals about what is going on in the research systems, these results have to be interpreted in relation to the complexity of the environment studied.

Within bibliometrics, two research streams have developed in parallel. Leaving their marks in the relevant bibliometric scientific literature (Borgman and Furner, 2002; Von Unbegrn-Sternberg, 2002):

- Von Unbegrn-Sternberg (2002) characterized one by a rigourous analysis of distribution properties that often result in statistical laws or mathematical models. This tradition is mainly based on the works of Lotka, Bradford and Zipf that became important bibliometric laws. *Bradford's law* suggested that only few journals include a proportionally large amount of the relevant articles in a subject field (Bradford, 1934). *Lotka's law* (Lotka, 1926) suggests that only a small number of authors in a subject field are highly productive. Price (1963) interpreted this law by saying that "half of the scientific publications have been produced by the square root of the total number of authors in a subject field". Finally, *Zipf's law* suggests that few words will appear many times in contrast with many words appearing only few times (Zipf, 1930). The three laws can be used to describe phenomena where there are a large number of common events, in contrast to a very small number of unusual events.

- The second tradition is more empirical, concentrating on studying relations extracted from scientific literature or from other bibliographical collections (Borgman and Furner, 2002). It is often based on citations, as in co-citation analysis or bibliographic coupling, or in words as co-word analysis. Some of the authors that have influenced and forged this bibliometric research tradition are Eugene Garfield with his works on citations (Garfield, 1955), Price's pioneering work on the networks of scientific papers (Price, 1965), Griffith (1988)'s influential discussions on citations, Narin *et al* (1984) with his work on patent evaluation and Henry Small's works on mapping co-citations (Small, 1973).

2.1.2 Co-authorship Index as an Indicator of Scientific Collaboration

The co-authorship index is simply defined as the average number of authors per document. Despite its simplicity, it has been considered as a way to document the scientific collaboration process between two or more authors (Bordons and Gomez, 2000; Borgman and Furner, 2002; Glänzel and Schubert, 2004). Although not much scientific attention has been given to co-authorship as compared to citation or co-citation indexes (Newman, 2004c), since the year 2000 co-authorship indexes have been widely used as proxies for the study of scientific collaboration (Bordon and Gomez, 2000). Hence, as Glänzel (2003) expresses, studying scientific collaboration through co-authorship measures has become one of the favourite topics in bibliometric research.

Notwithstanding, the co-authorship index is not exempt from criticism. In a broad sense, Bordons and Gomez (2000) define *scientific collaboration* as an activity that involves two or more researchers that as part of the same joint research project share intellectual, economic or physical resources. Consequently, scientific collaborations, similar to any activity involving human interactions, is a very complex process, from which the participation and contribution of each collaborator is very difficult to quantify (Bordons and Gomez, 2000). Moreover, Bordons and Gomez (2000) argue that as each collaborator may hold different aims and scope, different kinds of participation may occur. Authors' participation can vary from informally expressing opinions, exchanging ideas and data to working in constant interaction possibly varying at stages of the same project.

Katz and Martin (1997) and Laudel (2002) affirm that scientific collaboration involves more than the act of publishing a paper. As an example, published papers do acknowledge contributions from those other than the co-authors, however, co-authorship indexes fail to integrate these types of contributions as part of their calculations (Kochen, 1987). Moreover, Bordons and Gomez (2000) and Von Unbegrn-Sternberg (2000) draw attention to the fact that not all collaborations end with published works. Therefore, co-authorship indexes have to be used and treated with caution, having always in mind that they are measures related to only one of the stages in the scientific collaboration process. As a consequence, scholarly publishing indexes should be considered only as a partial measure of the complex scholarly communication system (Katz and Martin, 1997; Bordons and Gomez, 2000; Von Unbegrn-Sternberg, 2000; Laudel, 2002).

Despite criticisms, co-authorship is one of best documented, extensively indexed and easy to acquire metrics concerning scientific collaborations (Newman, 2004b). Moreover, despite their limitations, co-authorship indexes are invariant, ascertainable, non reactive, quantifiable and more importantly well documented; ample reasons to be widely used (Bordons and Gomez, 2000).

Collaboration patterns in bibliometrics have been studied at almost all levels.

- From an individual point of view to reveal co-operation patterns of individual scientists or to reveal aspects of their collaboration practices (Logan *et al*, 1991; Cronin, 2001).

- At aggregated levels to explore institutional collaboration patterns (Katz and Martin,1997; Laudel, 2002) or to reveal cross national cooperation (Katz, 1999; Calero *et al*, 2006; Havemann *et al*, 2006).
- At global levels to explore internationalisation in science (Frenken and Leydesforff, 2004; Glänzel and Schubert, 2004), or in an specific research domain such as the cases of Garcia-Ramon (2003) and Gutierrez and Lopez-Nieva (2001), who studied internationalization in geography and human geography journals respectively.

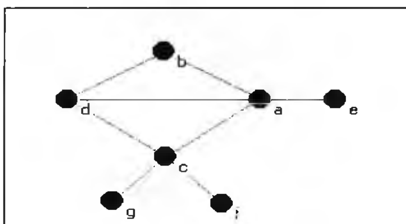
2.2 Representing Networks

In 1736, the work of Leonhard Euler led to the emergence of graph theory, a branch of mathematics that studies graphs composed of links and nodes and its applications. He did so, while solving what is now known the Königsberg bridges problem (Ramamurthi ,2004; Newman *et al*, 2006). The problem consisted of finding a path around a collection of bridges in 18[th] century Königsberg such as that its paths crossed each bridge exactly once. Referring to the city bridges' layout at that time, Euler proved that there was no solution to the problem as every island in Königsberg had an odd number of bridges coming out of it. Thus, the people of Königsberg could never go for a walk crossing each bridge only once (Ramamurthi, 2004). Since then, graph theory has been developed as a mathematical tool for describing the properties of networks and is considered today to be the basis of the study of networks (Borgatti, 2004). Formally, a *graph* can be defined as a set of vertices *V* together with a set of edges *E* that are mathematically denoted by:

$$G\ (V,E) \text{ or } G = (V,E)$$

Equation 2-1 Graph Definition

In its simple form, a network is a finite set of elements (nodes, vertices or points), linked by a finite set of connections (see left hand panel Figure 2-1).



node pairs	Geodesic Path
(a, e)	1
(c,b)	2 (via node a or d)
(b,f)	3

Figure 2-1 A graph representation of a simple undirected network (left hand panel). Right hand panel shows some distance measurements between pair of nodes.

Hence using the graph paradigm, elements and connections can be deployed to represent many kind of systems, such as people and sent emails (Koku and Wellman, 2002), computers and cables (Barabasi and Bonabeau, 2003), web pages and their links to other pages (Albert *et al*, 1999; Hayes, 2000), scientists and cited colleagues (Batty, 2003a), or scientists and their co-authors (Newman et al, 2000; Moody, 2004). Biological, social, technological or human networks have been or can be represented and modelled through what mathematically is defined as a graph.

Consequently, due to its simplicity and versatility, and taking advantage of newly available data sets, scholars have deployed formal mathematical representations of different kinds of networks based on graph theory concepts. At the same time, novel representations have been used as the base of devising new models that describe the behaviour (Newman *et al*, 2000) and the evolution of large complex systems (Barabasi *et al*, 2002). Hence, through the examination of network topologies, scholars have been able to better understand various large real-world phenomena. As a result, information networks such as the World Wide Web (WWW), where the elements are web pages connected by hyperlinks (Hayes, 2000), social networks where nodes are people connected by friendship, kinship or membership ties, or transportation networks where nodes are airports linked by airline routes, have all been modelled and examined.

2.2.1 Geodesic Path

The fundamental concepts borrowed from graph theory to analyze networks are mainly based on the notion of topological distance (Albert *et al*, 2002; Newman, 2003). In a graph, *distance* is defined as the number of links between a pair of nodes. In the case of Figure 2-1, the distance between node (a, e) equals 1 and that between nodes (a,g) equals 2 or 3, depending on via which node the connection is made. In an information network, distance can also be taken as a measure of node accessibility. Hence, short distances imply shorter paths to get from one node to another. Moreover, being easily reachable can determine who accesses first information, news, latest research or in general, any critical flow within the network. Additionally, it is easy to notice in Figure 2-1, that the number of indirect connections through neighbours helps to determine how accessible is a given node. For example, poor locally connected nodes such as *e, f* and *g* indirectly benefit from their few, but well-connected neighbouring nodes. Therefore, it is not only how well connected is a node, but also how well-connected its neighbours are.

Correspondingly, a *geodesic path* between a pair of nodes is defined as the shorter distance, chain of links, connections, steps or edges needed to reach one node from another (see right hand panel in Figure 2-1). In this context, all co-authors of a paper, all web pages listed within a given web page or the closest friends of a given person have a *geodesic distance* of one. In the same way, all web pages that need two clicks to be reached from a given page or the friends of the friends have a distance of two.

For an overall network, the *mean geodesic* averages the shortest node to node distance within the network's largest sub-community (Newman, 2003). Surprisingly, Watts and Strogatz (1998) reported that in a half million actor network, any two actors are separated through a small chain of over three actors on average. Moreover, Albert *et al* (1999) and Broder *et al* (2000), working separately and focusing on different parts of the WWW found that any two pages are only an average of 11 clicks from one another.

Network	No Nodes	Mean Geodesic	Size Largest Sub-Network	Diameter	Work Reference
Film Actors: actors linked by being cast together in the same movie	449,913	3.48	Around 90%	n/a	Watts and Strogatz . (1998)
One-day At&T Call Network where nodes are telephone numbers linked by calls made from one number to another	53,767,087	n/a	More than 80%	20	Abello *et al*, (1999).
WWW (nd.edu domain)	325,729	11.6	n/a	19	Albert *et al*, (1999).
WWW AltaVista : web documents linked by hyperlinks from one page to another	203,549,046	10.46	91%	16	Broder *et al*, (2000).

Table 2.1 Distance based properties of large-scale real world networks.

Some of the pioneering studies on networks of different kinds (Table 2.1) revealed that some large networks despite their sizes or nature possess relatively small mean geodesics.

2.2.2 Connected Sub-Communities

In graph theory, the *largest* or *one-component* corresponds to the largest interconnected sub-network, where at least one path can be found between any pair of nodes. However, its significance is adjusted according to the nature of the network under study. For example examining a co-authorship network, Horn *et al* (2004) presented the largest component as a relative index of the degree of integration within the discipline or academic field under study. In addition, within an information network such as the Internet, where nodes are computers and other communication devices connected by physical links (Albert and Barabasi, 2002), the largest component is taken as the largest fraction of the network within which communication is possible (Newman, 2003).

In general terms, one can say that the size of the largest component is a measure of closeness in non-purpose networks such as actors, citation, co-authorship or friendship, or of how effective a purpose-built network is at doing its job such in an airport or road system or in the case of the WWW, at making accessible web pages through hyperlinks. Broder *et al* (2000) reported that for the built WWW network, the largest component fills more than 91% of the nodes. Thus, a large number of pages are accessible from one another. Similarly, Watts and Strogatz (1998) reveal a highly interconnected actor network, where the largest sub-community comprises more than 90% of the participants.

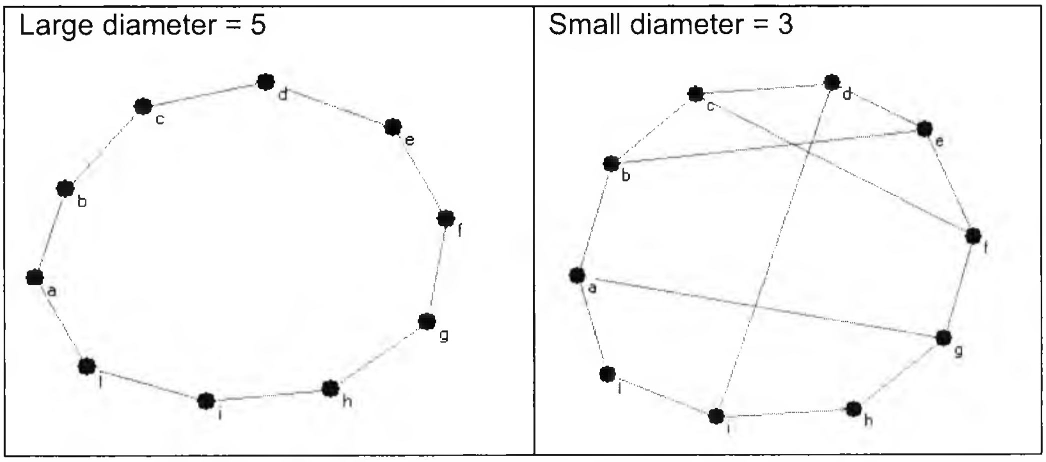


Figure 2-2 Comparing network diameters. Left hand panel shows a circular network with a large diameter. Right-hand panel shows the same network but exhibiting smaller diameter as the result of adding a few shortcuts.

Another important measure in the largest connected node set is *the diameter* that measures *the longest shortest path or the maximum geodesic* expected between nodes. In other words, it represents the maximum of the shortest distances between interconnected nodes. Thus, a large diameter (as shown in Figure 2-2 left-hand panel), especially in communication networks, is associated with delays because a message may have to pass through many intermediate points to get to its destination (Peterson, 1998).

However, adding only a couple of extra links (see right-hand panel Figure 2-2) creates shortcuts that make a smaller diameter, and therefore, a quicker communication process among nodes may take place. Results from studying the WWW at an individual level by Albert *et al* (1999) and Broder *et al* (2000) concluded that the maximum number of clicks expected for reaching a given web page (within the fraction of the WWW network analysed) are 19 or 28 respectively. Additionally, in a giant one-day call network built by Abello *et al* (1999), any telephone number in the largest component (see Table 2.1) can be linked to any other through a chain of no more than 20 calls. The studies listed in Table 2.1 agreed that very large real-world networks such as the Internet, the WWW or power grids, despite their large sizes, not only exhibited very short average distances between nodes, but also small diameters.

2.3 The New Science of Networks

As networks are everywhere (Newman *et al*, 2006), the study of what is called "*the new science of networks*" has gained increasing importance (Barabasi, 2002; Buchanan, 2002; Watts, 2003), and drawn the attention of the scientific community for the past ten years (Newman *et al*, 2006). According to Albert and Barabasi (2002) and Newman (2003), there have been at least two developments that prompted the increase in network research. First, the availability of large data sets about social, information, technological and biological networks. As a consequence, the research focus moved from analysing properties of individual nodes of small networks to analysing statistical properties of very large networks. Secondly, the availability of more powerful computational tools and the advent of new network communication technologies that allow the construction and study of networks with millions of nodes. Therefore, having the option of constructing networks that represent large biological, human-made, technological or social systems has made possible the understanding of not only the network structure, but also the study of how small localised changes in the topology may affect the whole system. However, different types of network structures have been proposed, modelled and evolved over time, according to specific characteristics of the systems that need to be described.

2.3.1 Ordered and Random Networks

Before the fifties, according to Albert and Barabasi (2002) and Newman *et al* (2006), most network research was focused on studying the properties of ordered networks. In an ordered network, like a crystal lattice, each node has the same number of edges joining the node and its neighbours in a tight cluster. Back then, graph theory was focused on ordered networks with no apparent design principles described as *random networks*, and first studied by Erdos and Renyi (1959). Through time, random networks have been seen as the simplest form to model a complex system, where all nodes have approximately the same number of links, thus the same probability of having new links.

Answering the question of whether or not real networks such as the WWW, the Internet, friendship, etc could be described as random networks prompted scientists to analyse these systems. In doing so, they discovered that most of these systems were far from random (Barabasi *et al*, 2002). Instead, large-scale networks such as Internet, WWW, friendship, power grids and natural neural nets were systems, from which it was not possible to attribute apparent generating principles (Barabasi and Bonabeau, 2003). Albert and Barabasi (2002) and Newman (2003) concluded that these networks had a mixture of randomness and free structure, and called them "*complex networks*". Moreover, a strong increase in the amount of research studying and modelling networks (Barabasi, 2002; Buchanan, 2002; Watts, 2003; Newman *et al*, 2006) suggested that the construction and analysis of networks' structures are important keys for understanding dissimilar complex systems such as those mentioned above.

2.3.2 Small-World Networks (SWN)

During the late fifties the sociological community started focusing on graph theory applications. Pool and Kochen(1978)'s paper, written during the fifties but published in the late seventies, which was the first work that raised basic questions about social networks. Instead of giving answers, the authors just posed questions such as how many people does a person know (*node's degree*)? Which is the statistical distribution followed by the degree measures? Which are the smallest and largest degree values? What kind of people have large degrees? How influential are those individuals in the network? and most important, What does the network structure look like? These were the questions that later helped to define the field of social networks, and those which the field has been dealing with since then. Besides, Pool and Kochen's (1978) untested mathematical model was the first to refer to the small world problem and from which Milgram (1967) based his own small world experiment. Milgram's aim was to find out how many acquaintances it would take to connect two randomly selected individuals. Consequently, he formulated what is named as the "small world problem" (SWP) that says "*Given any two people in the world, person X and person Z, how* *many intermediate acquaintance links are needed before X and Z are connected*" (Milgram, 1967, p.62).

Watts (2002) explains the experiment to solve the SWP as documenting and analysing how random selected subjects (passers) from a Nebraska telephone directory, will forward letters to a stockbroker (target) in Boston. The aim was, that using the knowledge of the structure of their own social networks, the participants will find out a person who was more likely to be close in any sense to the target subject, and so be likely to pass on the letter. According to Milgram's (1967) results, of the 160 chains that started in Nebraska, only 44 (27.5%) were completed (the rest were unsuccessful). On the completed chains, a median chain length of about 5.5 intermediate acquaintances was needed to pass the letter to the stockbroker; an unverified result (Kleinfeld, 2002; Watts, 2002) that Guare (1990) popularized as "Six Degree of Separation" in both his play in 1990, and a later film.

However, Kleinfeld (2002), Watts (2002), Newman (2001a) and Newman *et al* (2006) consistently argue that the results were taken as evidence of the SW hypothesis without any serious scientific scrutiny. The SW hypothesis states that most pairs of people in a population can be connected by a short chain of intermediate acquaintances, even when the size of the population is very large (Newman, 2001a). Moreover, Kleinfield (2000) on a closer examination of Milgram's experiment and results found an unsustainable gap between empirical results and Milgram's subsequent interpretations. Consequently, Kleinfield (2000) concludes that the SWP, as it was presented in Milgram's (1967) paper, did not have a reliable empirical basis. Nevertheless, a later paper on the SWP by Travers and Milgram (1969) sustained a much more detailed quantitative analysis and gave a clearer explanation about the SW experiments conducted by Milgram.

Watts and Strogatz (1998), in an attempt to model the SWP defined by Milgram (1967) found that in many real-world social networks nodes are connected by *relatively short paths* (number of links between nodes), despite the networks' sizes (Watts and Strogatz,1998; Watts, 2000). This property was termed the small world effect, but is more generally known as the *average path*. The *average path* measures the number of links that have to be included in order to travel from one node to another. Moreover, some observed networks had what is named as *high clustering* or *transitivity* (Newman *et al*, 2006), which in the language of social networks means that two people are much more likely to be acquainted with one another if they have a common acquaintance (Amaral *et al*, 2000; Newman, 2001a; Barabasi *et al*, 2002). In social network analysis (SNA), the *clustering coefficient* refers to the extent to which the existence of ties between actors A and B and between B and C implies a tie between A and C (Newman, 2001a). In doing so, the *clustering coefficient* accounts for the existence of cliques that represent circles of friends or acquaintances, in which, every member knows every other member.

On measuring these two properties, Watt and Strogazt (1998) defined a network to be SWN if it shows both *small average path length* and a *clustering coefficient higher* than one computed to a random network of the same size. Besides, they discovered that SWN interpolates between the two limiting cases of a high clustered regular lattice and a random graph with short path lengths between nodes (Ebel *at al*, 2003). For Barabasi *et al* (2003), the SW concept describes the fact that despite a complex system's large size, there is a relatively short path between any two nodes in most networks.

Performing empirical testing on the new SWN model, Watts and Strogatz (1998) discovered that many of real networks such as Hollywood movie actors , power grid and neural network of nematode *C. elegans* exhibit the characteristic clustering coefficient and path length presented by SWN. However, what their model failed to account for was the degree distribution of the nodes. For SWN, they assumed (but did not test) that the degree distribution had to follow a Poisson distribution as the random networks does (Watts, 2000). Besides, the SWN model did not take into account the mechanics for how new nodes and links are added, focusing only on static networks.

2.3.3 Scale Free Networks (SFN)

Trying to uncover the random nature of the WWW, Albert and Barabasi (1999) found that neither a random network nor an SWN may be used to totally model it. Rather than finding the WWW to be a random network as expected, they found that relatively few highly connected pages were essentially holding the WWW together. Hence, Albert and Barabasi (1999) introduced a new kind of network called scale-free networks (SFN). They advocated that SFN were the closest model of the topology of many complex systems where new links are not randomly added. Examples include the WWW (Adamic, 1999; Barabasi and Bonabeau, 2003), co-authorship networks (Newman, 2001a; Barabasi *et al*, 2002), power grid networks (Watts and Strogatz, 1998) and protein folding networks (Strogatz, 1998).

According to Barabasi and Bonabeau (2003), a network is a SFN if the distribution of its nodes' degrees (*degree distribution)* is spread according to a power law. Therefore, as shown in Figure 2-3, most nodes would have few connections, in contrast to a very few that would have a very large number of links. In that way, the power law distribution shows that not all nodes within a network have the same probability of being directly linked to the same number of nodes, as would occur in a random network.

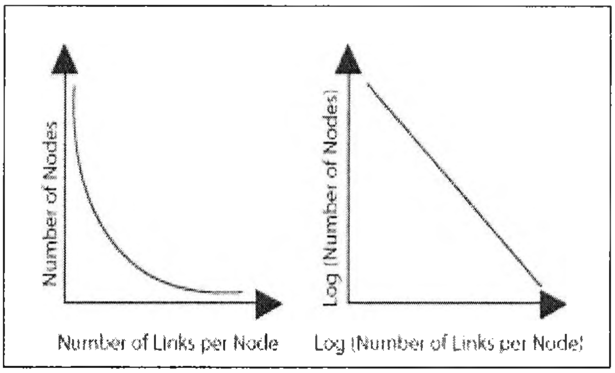


Figure 2-3 Power Law Degree Distributions (Barabasi and Bonabeau, 2003, pp. 53)

Additionally, Albert and Barabasi (1999) contended that SFN topology may be found in networks whose *growth* - extra nodes and new emerging links - is characterised by *preferential attachment* (previously known as the Matthew effect within sociology, Albert and Barabasi, 2003). Within a preferential attachment growth, they are the nodes with the higher number of links, those which are far more likely to be connected than others with fewer links. As a consequence, a special kind of node called a *hub* emerges. According to Barabasi and Bonabeau (2003), another important feature within an SFN is the density of the interconnectivity that creates two properties: robustness and vulnerability to attack. The former implies, as discussed by Albert and Barabasi (1999) and Newman *et al* (2006), that removing standard nodes will not easily lead to the breakdown of the network, due to the way nodes are connected. In contrast, simultaneous removal of the largest hubs will break down the network. The latter, vulnerability to attack, according to Barabasi and Bonabeau (2003), implies that due to the high network connectivity, viruses or epidemics can easily spread through the whole network, creating snow-ball effect failures. In other work, Barabasi and Albert (1999) found that the SFN structure is not a particular property of a special network such as the WWW. Instead, it can be found as generic properties of many large networks of very dissimilar nature like cellular metabolism, research collaborations, Hollywood actors, the Internet or protein regulatory networks, where the degree distributions follow a power law instead a Poisson distribution.

However, there is also some evidence that SFN may not be as widespread as they first seemed (Amaral *et al*, 2000). Amaral *et al* (2000) analyzed the degree distribution for a number of real networks. Their results pointed out that despite the fact that some of the analyzed distributions resembled power-laws such as the WWW or citations networks, others that at first sight seem to conform, actually do not.

According to Amaral *et al* (2000), the distributions of economical and technological networks such as a power electric grid in Southern California, the network of the world's airports and social networks such as the network of movie actor collaborations or the network of acquaintances of 43 Utah Mormons are far from exhibiting power law distributions. This study concluded that the essential limitation of SFN model is that it does not take into account the cost of a node getting more links. The SFN model defined by Albert and Barabasi (1999) assumes that a node can have as many connections, all cost-free, as it is able to accumulate and to maintain (Watts, 2000). Perhaps, this assumption may be valid for systems such as the WWW, where adding incoming or outgoing links to a web page does not involve any effort or cost. However, it can not be applied to systems such as power-grid, friendship or collaboration networks where adding a new cable between stations, finding a suitable colleague for co-authoring a paper or making a new friend involve costly processes of building, searching, discovering, and in the case of friendship also maintaining partners (Amaral *et al*, 2000; Watts, 2000). Therefore, in the same way that the SW model missed some features about real world, so did SFN models.

2.4 Social Networks (SN)

Social Network studies of all kinds have gained increasing importance and attention from the scientific community in recent years (Otte and Rousseau , 2002). Newman *et al* (2002) define a *social network* as a set of people or groups of people with some pattern of interactions or ties between them. In that way, friendships or business relationships among a group of individuals are some examples of social networks. Moreover, Newman (2001a), Barabasi and Bonabeau (2003), Glänzel and Schubert (2004) and De Nooy *et al* (2005) sustain the idea that the current society is a networked society, where almost everything is connected to everything else and where, as Koku and Wellman (2002) advocate, the group-centric organisations no longer fit the current society model.

As networks have become more and more ubiquitous (see global terrorists networks, the Internet, the WWW or online communities such as Myspace, Facebook, Wikipedia), Otte and Rousseau (2002) conclude that social network theory indirectly influences the way researchers nowadays think and formulate ideas on network structures that are very relevant to society. Moreover, scientists targeting networked systems of many kinds have found in Social Networks (SN) and Social Network Analysis (SNA), valuable tools for studying and investigating the social structures underlying their target systems.

2.4.1 SN Main Features

Otte and Rousseau (2002) emphasize that underlying any concrete network is a graph, thus, SN uses the graph paradigm to visualize, and some of its properties (see sections 2.2.1 and 2.2.2) to mathematically describe network structures. Hence, some graph concepts such as geodesic, size of the largest component and cliques have been taken as properties of the SN. Bühren (2002) summarizes the most significant network features (see table Table 2.2) that are used to measure specific network characteristics.

Characteristics of Relationships
Strength of the bond
Frequency of interaction
Number of resources which are exchanged by two network members in the relationship
Duration of the relationship
Symmetry and reciprocity of the exchange
Intimacy (emotional attachment).
Characteristics of Networks
Size or scope of the network
Density
Extent to which a network member is directly connected with others
Demarcation (share of all bonds of network members which are contained within the network)
Availability (average number of attachments which are necessary to connect the network members as a couple)
Homogeneity (extent to which the network members possess similar personal characteristics)
Cliques (network areas in which all members are directly connected)
Clusters (network areas with high density, but less stringently defined connection criteria than for cliques)
Components (network areas with which all members are directly or indirectly connected).

Table 2.2 Main features of SN (Bühren, 2002)

For example, Koku and Wellman(2002) showed positive results in analysing online learning communities examining concepts such as range (size and heterogeneity), density and boundedness, centrality, tie strength, multiplexity (multiple roles). Haythornthwaite and Wellman (1998) used work and friendship ties to examine their influenced on the type of media used for information exchange. With another approach, Glückler and Schrott (2003) used ego and socio-centric measures of network centrality (see Chapter 4) and range to measure actor centralities. They tested the impact of the network structure on leadership and team performance, as well as the difference between physical and electronic communication.

2.4.2 SNA: A Strategy for Studying Social Networks

Social Network Analysis (SNA), also referred as "Structural Analysis", is not a formal theory, but rather a broad strategy for investigating social structures (Otte and Rousseau, 2002). Its main difference to the traditional individualistic social theory is, as Mizruchi (1994) explains, that rather than isolate the individual behaviours of a social group, SNA advocates that actors take into consideration others' behaviours at the time to make choices. Thus, an actor's behaviour and decisions are influenced by other actors' behaviours, leading to a situation where relationships between actors become the first priority, and individual properties only are secondary (Newman, 2001a). In that way, Bührer (2002) advocates that SNA provides a framework to understand, the tools to visualize, and the language to talk about the interactions and processes of large-scale groups.

Moreover, SNA suggests that social behaviour and processes could be explained with reference to networks of social relations and the position of the actor within them (Mizruchi, 1994; Bührer, 2002; Otte and Rousseau, 2002). Thus, the tenet of SNA is that the structure of social relations determines the context of these relations (Mizruchi, 1994). For this reason, SNA is concerned with the social relations between interacting rather than isolated units of entities (Haythornthwaite and Wellman, 1998). For example, Mizruchi (1994) advocates that understanding the social relations among actors inside and outside their worlds would be helpful in explaining their worlds' governing laws. Both, Ahuja and Carley (1999) and Koku and Wellman (2002) pointed out, in the case of online communities, that viewing a community as comprising social relations enables analysts to examine the types of interactions that affect online communities.

Wheterell *et al* (1994) point out three important characteristics of SNA. Firstly, SNA conceptualises a social structure as a network with ties connecting members and channelling resources. Secondly, SNA focuses on the characteristic of the ties, rather than on characteristics of individual members. Finally, SNA views communities as "personal communities", that is, as networks of individual relations that people foster, maintain, and use in the course of their daily lives.

There are two main types of SNA: Ego Network Analysis (EGA) and Global Network Analysis (GNA). The former is related to the analysis of an individual's network such as the studies of Paul Erdős' collaboration network (De Castron and Grossman, 1990) or the Howard White description of Eugene Garfield's network (Otte and Rousseau, 2002). The latter type, GNA aims at finding all relationships between all network's participants in order to characterize the overall network structure. The analysis of the social network of economic organizations in Tampa Bay by Hagen *et al* (1997) or the study of the global terrorist network behind the 9/11 events by Krebs (2002) are examples of GNA. Otte and Rousseau (2002) reported an increase in the field's research, reflected in a steady growth in the number of publications in basic and applied SNA research. On the one hand, traditional SNA scientists are trying to improve some of the most used techniques such as network structural analysis or centrality measures. Among them, Borgatti (2005) reviewed the basic centrality measures and proposed new centrality measures attached not only to the type of trajectory that the traffic may follow (geodesic, paths, trails or walks), but also to the method used in the network to spread the information. Cornwell (2005) revised the definition of the largest component, as it assumes that all networks need to be interconnected, leaving out the isolated nodes that may play an important role. Thus, he proposed a set of centrality measures that use weights to incorporate the disconnected nodes.

But on the other hand, despite the real growth of the SNA field started in the 80s, Otte and Rousseau (2002) also reveal a moderate growth in publications in databases from other disciplines outside sociology such as Medline Advanced and PsyInfo. It seems that scientists from other disciplines rather than sociology have started using quantitative methods from sociology such as SNA for analyzing networks, not only social but of many kinds.

2.4.3 New Approaches to Understanding Social Networks

In the past, traditional studies of social networks were mainly applied to examine very small or medium static social networks (Newman, 2001a), and were mainly focused on the network's statistical properties rather than on the network structure per se (Newman *et al*, 2006). In contrast, the new science of networks, according to Barabasi

(2002), is concerned not only with mathematical properties of the networks, but also with studying the structure and evolution of networks that grow unplanned and decentralized. Moreover, the advent of the Internet, cheaper computer power and large amounts of available data have changed substantially the dimension of how graph theory, bibliometrics and SNA can be all part of the same powerful approach to study large-scale networks.

Thus, since around the year 2000 the tendency to study networks is by applying a highly multidisciplinary approach, whose theoretical background involves concepts from social networks, graph theory, SNA and in the specific case of collaboration networks, calculating bibliometric measures such as co-authorship indexes. In this way, the characterization of entire disciplines as in the case of sociology by Moody (2004), the visualization of the evolution of citation and co-authorship networks by Börner *et al* (2004) or mapping global science by Wagner and Leydesdorff (2003) have become achievable.

2.4.3.1 Co-authorship Networks

Derek J. De Solla Price is considered one of the pioneers on understanding the nature of scientific research (Newman *et al*, 2006) through the construction of a network of papers (nodes) linked by citation (Price,1965). Despite having very limited computing resources, Price (1965) was one of the first to represent citation connections as a network and to analyze the basic network's properties.

Analysing co-authorship networks is nothing new in mathematics (Grossman and Ion, 1995; Batagelj and Mrvar, 2000). Within the mathematical scientific community, Batagelj and Mrvar (2000) explain there is an important figure that measures the distance of any mathematician to Paul Erdős, who is thought to be the centre of the math collaboration network. Paul Erdős was an unusually prolific Hungarian mathematician who died in 1996. The Erdos' collaboration graph, by Batagelj and Mrvar (2000), is comprised of 507 co-authors that collaborated with Erdős in writing approximately 1500 papers. In total, it had 6100 nodes and 9939 edges. The Erdős number is calculated according to the distance between a given mathematician and Erdős himself. Thus, everybody who wrote a paper with him has an Erdős' number of one. Those who collaborated with those who have an Erdős number one, have an Erdős' number two and so on. Though mathematicians are regarded as less collaborative scholars, an aspect shown independently by Albert Barabasi (2000) and Newman (2001a), the Erdős collaboration network demonstrates that they form a very highly interconnected network. As Watts (2000) and Barabasi (2002) agree, the very existence of the Erdös' number shows the presence of a significant community structure behind a co-authorship network.

2.4.3.2 New approach to Co-authorship Networks

As previously mentioned, early bibliometric work on scientific collaborations were mainly focused on, as Börner *et al* (2004) state, describing major statistical features of static bibliographic data sets. However, the progressive availability of comprehensive digital repositories concerning collaborative scientific outputs such as scientific papers, pre-prints, grants, patents or conferences proceedings, coupled with the increase in computer power, have propelled a progressive interest towards understanding not only from Scientometrics scholars but other scientists as well.

The way that collaboration or co-authorship networks were investigated changed back in 2000 when Mark Newman presented his findings about the structure of scientific collaborations of computer science, biomedical research, theoretical physics and high-energy physics from 1995-1999. Newman's novel approach consisted of taking advantage of the large volume of scientific literature indexed by digital bibliographic repositories such as the ISI-Web of Knowledge (ISI-WoK) and Medline to model co-authorship networks behind any discipline. Using the co-authorship network as a proxy for SN, he deployed standard concepts from SNA to investigate, and as well as less standard metrics such as path length or clustering coefficient, to analyse the network properties. Despite the network sizes, he found high clustering coefficients and short average path lengths between the scientists in the analyzed networks. The results showed collaboration networks with structures far from random, but in accordance with the SW structures described by Watts and Strogatz (1998).

After Newman's work, many others using similar approaches have concentrated efforts to uncover collaboration network structures at different levels. Moody (2004) analysed social integration within social sciences, Liu *et al* (2005) examined the state of the co-authorship networks in the digital library research community for the last decade. Nascimiento *et al* (2003) investigated collaboration links from SIGMOD conference papers, Horn *et al* (2004) described the evolution and impact of research in computer-supported cooperative work, and Wagner and Leydesdorff (2005) attempted to map the global collaboration networks in science.

Barabasi *et al* (2002) proposed a simple model that captures the evolution of collaboration networks over time. They argue that evolution within a collaboration network follows "*the richer get richer*" principle, where scholars with a higher number of collaborators are more likely to collaborate than authors with few co-authors (see 2.3.3 sections on SFN). Working also on dynamic networks, Börner *et al* (2004) created a model that simultaneously evolves multiple types of networks. The model involves a collaboration network where authors are linked through co-authorship, and a citation network where papers are linked by citation.

The main difference between the two pioneering works from Newman and Barabasi was that Barabasi's team concentrated on the network's evolution, rather than static properties of the network, as was Newman's study. Nevertheless, both studies were the first to point out that collaboration networks do not tend to show random structures rather they should be modelled as SFN or SW. In that way, not all scientists have the same number of connections, however, only few of them may exhibit a disproportionate number of co-authors or publications, corroborating what Lotka established regarding the distribution of scientific productivity as early as in 1926.

3 Building Scientific Collaboration Networks - The Special Case of GIScience

Since early 2000, a new approach to building collaboration networks that stands at the crossroad between bibliometrics, social network analysis, graph theory and statistical mechanics, introduced conceptual tools that have facilitated the representation, characterization and evolution of entire disciplines. Consequently, scientists from diverse backgrounds have been able to map larger scientific networks built from collaborations intrinsically recorded in papers' co-authorships. In that way, studies of scientific fields such as Physics (Newman, 2001a), Mathematics and Neuroscience (Barabasi *et al,* 2002; Newman, 2004c), Social Science (Moody, 2004), and Global Science Networks (Glänzel and Schubert, 2004; Boyack *et al,* 2005) have outlined patterns of collaboration, mapped and located the scientific interrelations and generally offered revealing insights into the social structure of these fields.

Based on the same set of concepts, this study aims to describe GIScience collaboration networks, tailored to the intrinsic characteristics of GIScience as an emergent multidisciplinary field, which addresses fundamental and applied research.

3.1 Scientific Disciplines

The term *discipline* has being used to describe an organizational unit in both, educational programs and knowledge production (Pierce, 1991). Specifically, while the former refers to the ordering of knowledge for teaching purposes in academic institutions at any level (Stichweh, 2001), the latter considers a discipline as a branch of instruction for the transmission of scientific knowledge (Dogan, 2001).

3.1.1 Knowledge Production Models

Traditionally, the disciplines have been very controlling in the organization of the science system (van den Besselaar and Heimeriks, 2001), and also in the establishment of rigid reward systems and parameters for indirectly managing the careers of those associated (Whitley, 2000) up to the point that, as Gibbons *et al* (1994) critically contend, social characteristics of knowledge production have been long established by disciplinary sciences such as chemistry, physics and biology.

Gibbons *et al* (1994) identify as the "old paradigm" or Mode 1 of scientific discovery, the one that focuses on the production of theoretical or pure knowledge, mainly adopted in theoretical and experimental sciences. Moreover, Mode 1 establishes a clear distinction, as Albert (2002) explains, between basic research and research aimed at solving problems. Within Mode 1, knowledge production centres are frequently Universities organized into faculties and departments with relatively stable hierarchies in terms of organizational structures and practitioners (Nowotny *et al*, 2003). Hierarchies, which at the same time, are in accordance with existing disciplinary lines (van den Besselaar and Heimeriks, 2001), and their respective subject domains. Consequently, each discipline has developed and established norms and criteria for what counts as authoritative knowledge (Pierce, 1991), which is mainly validated through the peer-review processes (Whitley, 2001). However, the new dynamics of science (Nowotny, *et al*, 2003; Morillo *et al*, 2003; Leydesdorff, 2006a Leydesdorff, 2006b), new attitudes toward the social relevance of research results (van den Besselaar and Heimeriks, 2001) and the fact that one discipline cannot accumulate all relevant knowledge (Nowotny, *et al*, 2003) have led to a transformation of the processes of research and production of knowledge (Gibbons, *et al*, 1994; van den Besselaar and Heimeriks, 2001; Nowotny *et al*, 2003; Whitley, 2000). Therefore, it is undeniable that the nature of the research process is being transformed (Nowotny *et al*, 2003), consequently, a new mode of application-oriented research has been emerging in parallel with traditional academic research (Gibbons *et al*, 1994).

Gibbons *et al* (1994) proposed this new knowledge production model as Mode 2 that contrary to the traditional Mode 1, does not function within disciplinary frameworks, and is the result of transdisciplinary or cross disciplinary research carried out by scholars from diverse backgrounds. Transdisciplinary knowledge production in Mode 2 exhibits a constant interaction between fundamental and applied research. Thus, it does not make a distinction between a theoretical core and applied sciences, such as frequently seen in engineering. Therefore, Gibbons *et al* (1994) and Nowotny *et al* (2003) advocate that Mode 2 creates a model where knowledge easily flows across disciplinary boundaries, where a mobile human force and research organization is more open and flexible.

Nonetheless, Albert (2002) agrees that knowledge production is undergoing transformations. He draws attention to the fact that those transformations vary considerably according to areas of research, disciplines and institutions. Therefore, it is very unlikely that the two knowledge production models fit each discipline's specific characteristics perfectly, as interdisciplinary embraced in different ways within different disciplines. There are disciplines for which, interdisciplinary is a means of augmenting their theoretical and empirical frameworks via new questioning (Albert, 2001). Albert (2001) explains that to some extent various disciplines consider interdisciplinary as a direct link to what is named as "semi-scientific" knowledge. Thus, knowledge production needs a thorough examination within the context of each discipline's theoretical or applied research framework.

3.1.2 GIScience as a Scientific Field

Broadly, GIScience is an inherently multidisciplinary field underpinning GISystems (Mark, 1999). Hence, Geographical Information (GI) and Geographical Information Systems (GISystems) are two elements that cannot be detached from the establishment and evolution of GIScience as research domain. As Mark (1999) explains, spatial data challenges and new applications motivate the science. At the same time, advances in GIScience fundamental issues enhance knowledge of geographical concepts and theories that improve applications. Likewise, the GIScience research community has made the limitations of GI and GISystems an integral part of its research (Goodchild, 1991).

3.1.2.1 GI, GISystems and GIScience

Geographic Information (GI), a central part of GIScience, consists of facts about specific places on the Earth's surface (UCGIS,1996). As "*almost everything that happens, happens somewhere. Knowing where something happens is critically important*". (Longley *et al*, 2001, p.2), Mark (1999) points out that the appropriated usage of GI implies great potential to extend the capabilities of scientific researchers, decision-makers and the general public. Accordingly, over the last four decades problems within society that involve any aspects of location (GI), and the limitations of GISystems in handling it, have prompted extensive scientific attention (UCGIS, 1996; Longley *et al*, 2001; Mark, 2003).

The origin of GIScience as a research field cannot be detached from history and evolution of GISystems. Their sharing of the word *Geographical* not only denotes the importance of and their mutual dependence on GI, but also the independence on one another (Fisher, 1998). Longley *et al* (2001) state GI is central to the appropriateness of GISystems for representing, analyzing and tracking spatial activities, which make GISystems stand out from traditional information systems. On the other hand, the distinctive and unique nature of GI (Goodchild, 1992; Kennedy, 1994) has formed a loose interdisciplinary research community under the GIScience domain, challenged by a coherent set of research questions and problems concerning GI handling, representation, modeling, analysis, storing, sharing, distribution, access, visualization and ontological issues (Goodchild, 1992; UCGIS, 1996; Mark, 1999; Mark, 2003; Goodchild, 2006).

GISystems' early developments were led by the implementation of land management and governmental applications (Mark, 1999). At that time, the aim of developing and using GISystems was to fulfil an existing technological gap regarding the manipulation of spatial information (Goodchild, 1992). However, over time as GISystems have become ubiquitous tools that fall somewhat in the middle of any discipline dealing with the distribution of phenomena on the surface of the earth (Wright *et al* , 1997), its limitations and challenges have also increased. From the 1980s onwards, the idea that research on the advancement of GISystems technology may have enough substance to be considered as an academic field of study surged (Mark, 2003). Nonetheless, it ignited a lengthy debate between its supporters (Goodchild, 1991; Openshaw, 1991) that perceive GISystems as almost of a saviour (Wright *et al*, 1997), and its critics, who could not conceive the idea of a discipline behind a technological tool (Wright *et al*, 1997; Schuurman, 2000) that under their eyes lacked a real knowledge production system (Taylor, 1990).

To demystify the GIScience-GISystems relationship, Wright *et al* (1997) suggest a continuum model where each part is placed at opposite extremes. At one side, GIScience is concerned with academic areas and their members related to the development of theories and methods. At the other side, GISystems are used as a software tools (Fisher and Unwin, 2005), intermediating between the science and the system (Wright *et al*, 1997). Fisher and Unwin (2005) perceive Wright *et al* (1997)'s polarization theoretically problematic. Their continuum model implies that users of GISystems may only engage within any fundamental discussion on GIScience in their own scientific domain, thus, leaving only GIScience researchers the task of doing valid spatial science (Fisher and Unwin, 2005). GISystems users will always be tool users, but never authoritative GIScience scholars because they would never be able to validly scrutinise, criticize or re-formulate any GIScience principles. Thus, Fisher and Unwin (2005) argue that Wright *et al* (1997)'s model appears notably inadequate, because the uncritical use of any system would never lead to good science.

Fisher (1998) proposes an alternative model that represents the GIScience-GISystems relationship as a cyclic interaction, in which theory can, at the same time, underlie and be used to produce the tool. Thus, new representations and concepts related to

GIScience may be implemented, if applicable, as part of any GISystems. The GISystems-GIScience cycle is conceived as an open system where new concepts from allied disciplines are welcome and legitimated. Fisher's (1998) model makes evident the important role that GISystems and GIScience have played in linking disjointed fields such as photogrammetry, remote sensing, geodesy, cartography, surveying, geography, computer science and spatial analysis (Mark, 2003), thus, acknowledging the multidisciplinary nature of the GIScience research domain (Mark, 1999; Cova, 2000). It clearly separates the science from the systems without completely isolating one from another.

3.1.2.2 GIScience as an Academic Domain

To be taken seriously as an academic domain, the GIScience research community has been demonstrating that despite the reciprocal connection between the technology and the science, the domain possesses an internal coherence and a unique set of fundamental questions (Goodchild, 1992) that go beyond the technology. Wright *et al* (1997) present some conditions to establish the emergence of a science out of a technology that authors agree GIScience met (Mark, 1999; Goodchild, 2006). Firstly, GISystems is a driven technology behind the science from which the challenging issues behind its developments cannot be placed under any existing disciplines.

Moreover, the GI applicability notion has been exported to many other disciplines where it has been found to be useful. GISystems have passed from being exclusive tools for expert users or scholars working on environmental or geosciences, to pervade many other research disciplines and society in general (Cova, 2000). Nowadays, there is no lack of evidence of the relevance of GISystems as a technology for modern society (Mark, 2003; Goodchild, 2004; Fisher and Unwin, 2005; Goodchild, 2006).

Despite some critics arguing that GIScience was just repeating axioms from other sciences and technologies (Taylor, 1990; Schuurman, 2000), Openshaw (1991) emphasized that the uniqueness of the technologies encompassed in GISystems had exceeded the boundaries of geography and other disciplines such as computer science (Goodchild, 1992). For example, GISystems overcome the lack of capabilities of traditional information systems to handle GI, a fact that makes then valuable software tools. Moreover, dual keys (such as *x,y* or latitude, longitude) are not unique to GI (Goodchild, 1992), but traditional query systems did not include basic methods to handle them, as spatial keys are based on two continuous dimensions and can be associated with a *z* value (topographic elevation); GISystems are able to represent continuous surfaces (Longley *et al*, 2001) with an unlimited infinite number of tuples <*x,y,z*>. Therefore, these generate large files (Goodchild, 1992; Longley *et al*, 2001) that require time and resource consuming processes and special tools for integrating different types of GI (UCGIS,1996). Features that are not present in traditional information tools. Moreover, traditional information systems did not offer appropriates conceptual data models to represent geographical phenomena. Finally, GI measures observe *spatial dependence* where "*everything is related to everything else, but near things are more related than distant things*" (Tobler, 1970; Tobler, 2004). Likewise, spatial dependence methods implemented as part of GISystems have allowed the characterization of complex geographic landscapes (Goodchild, 2004).

However, GI do not only possess unique features, but also have posed far-reaching challenges such as the development of theoretical frameworks to properly handle time and the third dimension (Goodchild, 1992; Longley *et al*, 2001), overcoming the limited representation of geographical phenomenon on a flat surface rather than on a real curved surface (Wright *et al*, 1997), which are of enough scientific significance and complexity that cannot be placed under any of the existing disciplines (Goodchild, 1992) and truly justified the existence of the discipline in its own right.

3.1.2.3 GIScience Definitions

In 1984, the growing of significant funding by the US National Science Foundation (NSF) for the National Center for Geographic Information Analysis (NCGIA), considered as a milestone within the development of the discipline, drew the attention of the GISystems academic community towards the science rather than the tool (Mark, 2003). The progressive establishment of the field as a science shifted the research community's focus away from the tool towards the analysis of more fundamental issues raised by the use of GI and GISystems (Goodchild, 1992; Wright *et al* 1997; Mark, 1999). Hence, a transition from being a community of people with an interest in a software tool (Wright *et al*, 1997) towards a research community more focused on answering questions of fundamental scientific significance surrounding GISystems (Goodchild, 1992) started. Consequently, the change from *System*s to *Science* (Goodchild, 1992) as part of the field's name reflects the evolution of the research community towards a serious discipline than presents, as Goodchild (1992) asserted, a range of intellectual and scientific challenges spanning many disciplines and fields that surpass what GI handling implies.

Since 1992 when Goodchild first used the term *Geographic Information Science* and justified the academic nature of the field, there have been different attempts to define the discipline (Goodchild, 1991; UCGIS, 1996; Wright *et al*, 1997; Longley *et al*, 2001; Cova, 2002; Mark, 2003; Goodchild, 2004), many of them influenced by the technological context and the focuses of GIScience' research agenda at that precise time. In fact, Goodchild (1992) avoided giving a formal definition of the discipline; instead, he argues that research in GIScience aimed at overcoming the limitations surrounding GISystems technologies that were hindering its successful implementation, and also, at understanding the issues related to the technology's potentials. In a similar tone, a UCGIS (1996) definition suggests that GIScience "*has emerged as an acceptable umbrella term for the fundamental problems surrounding the effective capture, interpretation, storage, analysis and communication of geographic information*". Wright *et al*, (1997) defined GIScience "*As the science of GIS, concerned with geographic concepts, the primitive elements used to describe, analyze model, reason about, and make decisions on phenomena distributed on the surface of the earth* ". Later in 1999, a full definition of GIScience was presented at the NSF.

"Geographic Information Science (GIScience) is the basic research field that seeks to redefine geographic concepts and their use in the context of geographic information systems (GIS). GIScience also examines the impacts of GIS on individuals and society, and the influences of society on GIS. GIScience re-examines some of the most fundamental themes in traditional spatially oriented fields such as geography, cartography and geodesy, while incorporating more recent developments in cognitive and information science", (Mark, 1999, p. 2).

The NSF's definition draws attention to the GIScience-society relationship and also, to the interdisciplinary nature of a field that embraces relevant developments from other disciplines, and, focuses on solving fundamental issues regarding GI within traditional spatially-influenced disciplines. Later, Goodchild (2004) regarded the discipline as "*a branch of information science that deals with places on or near the surface of the Earth*" and alternatively as "*the set of fundamental principles underlying the design, testing, and use of geographical information technologies*" (Goodchild, 2004, p.1). The two definitions cover two basic aspects of GIScience research, the former focuses on solving fundamental issues concerning all aspects of GI, and the latter on how to use its fundamental findings to advance GI technologies.

All previous definitions convey an initially strong focus on GISystems as the main problems to solve were related to data handling, a consequence of restrictions of hardware and software specifications of that time. Equally, a close relationship existed with geography, the traditional academic field that studies the Earth. However, later definitions make clear that GIScience's research agenda is focused around a more elaborate set of questions, and challenges concerning GI and Geographical Information Technologies (GIT) such as GPS, remote sensing, image processing, soft photogrammetry and virtual environments (Cova, 200; Mark, 2003; Goodchild *et al*, 1999; Goodchild, 2004) that have led to the development of two research streams, basic and applied that cannot be divorced one from another. Mark(1999) expresses that GIScience is clearly an area in which applications motivate the science, but also, awareness of the theory may help to improve applications. If GIScience scholars aim to advance the discipline and to move it away from being an application-driven field to a discipline, the right balance between applied and pure or theoretical research needs to exist (Mark, 1999).

3.1.2.4 GIScience Research Agenda

In 1995, a number of research groups from US universities, national laboratories and learned societies formed the University Consortium for Geographic Information Science (UCGIS). Since then, the UCGIS has focused its attention on the identification of research priorities and challenges that aim to move the discipline forward. In 1996, the UCGIS presented the first GIScience research agenda that comprised topics concerned with interoperability, data handling and distribution, new geographical representations, cognition, GIS and society, scale, the future of the spatial information infrastructure and uncertainty. This early agenda reflected the discipline's challenges mainly associated with hardware limitations and the incipient data distribution systems of the time, and also with the scarce knowledge of the GIS-society relationship. In contrast, the latest research agenda has focused on short and long term goals for the discipline driven by a deeper concern about how society interacts with GI and GITechnologies. The agenda comprised short term research priorities addressing the limitations on the implementation of the newest GITechnologies, and long term research priorities focused on spatial ontologies, geographic representations, spatial data acquisition and integration, remotely acquired data and information in GIScience, scale, spatial cognition, space/ time analysis modelling, uncertainty, visualization, GIS and society, geographic information and engineering.

3.1.3 Knowledge Production and GIScience

As science and researchers have become more specialized, and posed more complex research questions, knowledge concerned with other subfields or with completely different disciplines is frequently needed (van den Besselaar and Heimeriks, 2001; Morillo *et al*, 2003), resulting in new interdisciplinary approaches for knowledge production (Gibbons *et al*, 1994; Morillo *et al*, 2003; Leydesford, 2006).

It is no different within a GIScience setting, as fundamental questions need to be addressed by researchers working in different existing disciplines, and often conducted within very different research traditions (Goodchild *et al*, 1999; Mark, 2003). Goodchild *et al* (1999) broadly grouped the possible components of GIScience research and its allied disciplines into four categories.

1. Research on geographic information technologies (GIT) such as cartography, remote sensing and image processing, geodesy, surveying, photogrammetry and mobile computing.
2. Research on digital technologies and information in general, such as computer and information science.
3. Research on the Earth (particularly the physical or human aspects) such as geology, geophysics, oceanography, agriculture, biology, environmental science, geography, sociology, political science, anthropology, among many.
4. Study the nature of human understanding, and its interactions with machines such as cognitive psychology, environmental psychology, cognitive science and artificial intelligence.

In accordance, Mark (1999, 2003) advocates in his definition of the discipline that GIScience research plays a double role of re-examining fundamental concepts of other disciplines (mainly traditional geosciences), but also, drawing concepts from other disciplines (Mark, 1999; Goodchild, 1999). Therefore, today it is recognized that GIScience cannot rely upon uni-disciplinary solutions for most of the challenging problems posed by GI and GI Technologies (Mark, 1999; Godchild, 1999).

In contrast to traditional Geography, in which knowledge creation is primarily the preserve of universities and follows Mode 1 (refer to section 3.1.1), the seeking of knowledge in GIScience has Mode 2 characteristics (Longley *et al*, 2001), where research is context specific and multidisciplinary rather than pure and discipline based

(Gibbons *et al*, 1994). Longley *et al* (2001) present the following list of characteristics that suggest GIScience follows mode2 of knowledge production.

- The presence of applied and basic research on GIScience (Goodchild, 1992; Mark, 1999) that makes GIScience research draw on fundamental laws, instead of formulating them.
- Multidisciplinary team work that according to UCGIS (1996) is mainly the norm.
- The lack of universal criteria to measure research success in GIScience. According to Longley *et al* (2001), academic research success in GIScience is measured through the number of publications in referee journals. However, other areas that deal with applied GIScience and GISystems such as industrial, governmental or other users may have different criteria to measure success. Therefore, Longley *et al* (2001) conclude refereed journals are not the sole forums to publish research results or findings.
- GIScience's Scholars are perceived as having multidisciplinary backgrounds (Mark, 2003).

The definitions, fundamental characteristics and aims of the research agenda generally portray GIScience as a multi and cross disciplinary discipline (Goodchild, 1992; Goodchild, 2005). Additional, some definitions present GIScience as inter-disciplinary discipline with some connotations of a trans-disciplinary domain (Mark, 1999; Cova, 2000; Goodchild, 2004). However, comparing the definitions of these 3 concepts (multidisciplinary, inter-disciplinary and trans-disciplinary) it seems that these terms have been used interchangeably, even though, as van den Besselaar and Heimeriks, 200 explain, they are not synonyms. To what extent is GIScience an interdisciplinary domain is a difficult question to answer. However, to be interdisciplinary the research approach used by GIScience, not only needs to examine the problems from the angle of different disciplines, but also all disciplinary views of approaching the problems have to be integrated as a coherent methodology. As Goodchild *et al*, (1999) discuss, there is a concern in the field on the emergence of a non-systematic research framework, as a consequence of the number of isolated efforts that have been missed in between dissimilar research environments. Mark (1999) describes GIScience as "an emerging cross-disciplinary field", using a near synonym of interdisciplinary. However, the current reality shows an interdisciplinary field only in the scope (Goodchild, 2006).

3.2 A New Approach for Building Collaboration Networks in GIScience

Building collaboration networks from highly multidisciplinary disciplines such as GIScience poses a new set of challenges if compared to building co-authorship networks from well-established or theoretical disciplines. On the one hand, traditional disciplines tend to have a clear-cut set of journals (Morillo *et al*, 2003) that are indexed under subject categories corresponding to disciplines in the bibliographic databases. As analyzed in the previous section, research in GIScience is a combination of basic and applied problem-solving methods that lead to a very unclear set of potential journals as publishing sources. Therefore, an alternative strategy was designed for selecting GIScience's relevant publishing outlets, from which a network representing the discipline's scientific collaborations can be constructed.

3.2.1 Delineation of Disciplines from Academic Journals

Within a discipline, as Pierce (1991) describes, research is done based on uniform training and through sharing a series of common information services. These formal scientific communication systems, which vary from discipline to discipline, comprise scientific or technological journals, conferences, workshops or academic books. Traditional disciplines, as van den Besselaar and Heimeriks (2001) explain, have been dominant in the organization of science, but also in establishing reward and career systems of their associates. Thus, within the intellectual and social structure of each discipline, forms of internal scientific communication can be seen as instruments for filtering and communicating findings (Morillo, *et al*, 2003) that reward successful research strategies. Therefore, as Whitley (2000) argues, communication forums are used not only as a means of convincing fellow researchers of the importance and relevance of the results, but also as a way to enhance one's own reputation within a scientific community.

Consequently, to organize disciplines regarding scientific literature, they have been operationalized in terms of journal sets (Leydesdorff and Cozzens, 1993). Derek de Solla Price (1963) speculated that disciplines would commence displaying some degree of specialization when the carrying community grew larger than a 100 or so active scholars (Crane, 1972; Kochen, 1983). In this respect, according to Stichweh (2001), the emergence of a scientific field or sub-field is equivalent to the invention of new communication forms tailored for the carried disciplinary communities, hence, new journals and conferences would emerge too.

The categorization of academic journals for organizing scientific literature, made in the early 1970s and expanded in 1981, is the result of a subjective assignment and a cross-examination of citation patterns among disciplines (Leydesdorff and Cozzens, 1993). It is this classification of journals into subject categories that specialized data bases such as PNAS (*Proceeding of National Academy of Sciences of United States of America*), Medline and IEEE-Xplore, or those with wider scope such as the Web of Knowledge (WoK) by the Institute for Scientific Information (ISI), follow when classifying the pertinent scientific literature. However, the evolutionary and dynamic structure of the current sciences (Leydesdorff, 2006), to some extent augmented by an increase in the amount of academic cross-disciplinary interrelations among scholars carrying out applied research (Gibbs *et al*, 1994; van den Besselaar and Heimeriks, 2001; Morillo *et al*, 2003), present a problem for the evaluation of science schemes that use these static categories (Leydesdorff and Cozzens, 1993; Klavans and Boyack, 2006). Moreover, despite the fact that for example ISI's databases ad subject categories are subject to periodic updates (Morillo *et al*, 2003), it is a demanding task to keep up with the pace of current science, which is marked by the constant emergence of new disciplines and the division of fields into new sub-fields (Leydesdorff, 2006b). Moreover, one cannot be sure that new emerging disciplines may statically organize and delineate around fixed journal sets (Klavans and Boyack , 2006).

In order to represent more accurately the structures of current disciplines, especially those identified as highly interdisciplinary such as biotechnology, bioengineering or environmental sciences (Leydesdorff, 2006a), authors have devised a series of dynamic journal *relatedness measures* that evaluate journals closeness using mainly bibliometric or social network measures such as citation as co-citation (Morillo *et al*, 2003; Klavans and Boyack, 2006; Leydesdorff and Cozzens, 1993; Leydesdorff, 2006b) and betweenneess or closeness centrality respectively (Leydesdorff, 2006; Leydesdorff, 2006a). In doing so, the overall aim is that journals within the same disciplines or sub discipline can be classified all together in the same category (Boyack *et al*, 2005).

Leydesdorff and Cozzens (1993) propose an analytical approach based on bibliometrics to dynamically attribute journals to specialities. In doing so, they extract journal-journal citation from the Science Citation Index (SCI) and applied multi-variate analysis comparing loading factors to detect any systematic changes or clusters in the fields' structures. Morillo *et al,* (2003) present an approach to define a disciplines' topology using journal-multiassignation to ISI categories. The authors assume that journals appearing in more than one category should be more interdisciplinary than those single-assigned (Morillo *et al*, 2003). From social network measures, Leydesdorff (2006) deploys betweenness centrality as a means of identifying in-between journals within journal-journal citation environments. The author assumes that journals with the highest betweenness measures can be considered as proxies for specialities and disciplines (Leydesdorff, 2006). Finally, Leydesdorff (2006a) proposes another dynamic approach to identify interdisciplinary environments based on a journal-journal citation matrix through the identification of the overlapping set of journals between SCI and Social Science Citation Index (SSCI). Using betweenness measures and vector space visualizations to map the citation environment, the author suggests that closer journals are more disciplinarily related than distant marginal journals.

3.2.2 GIScience in Terms of Academic Journals

To some extent, the process of selecting data sources for building collaboration networks from well established disciplines has proved less complex (Newman, 2001a; Barabasi *et al*, 2002; Moody, 2004) as pertinent journals are mainly indexed under clear-cut subject categories in bibliographic databases (Morillo *et al*, 2003).

In contrast, as GIScience's knowledge production exhibits mode 2 characteristics (Longley *et al*, 2001), in which basic and applied research are carried out by multidisciplinary teams deploying interdisciplinary approaches, relevant results and findings are frequently published in journals from different disciplines. Consequently, its multidisciplinary nature and its interdisciplinary scope make GIScience a field with very fuzzy and dynamic disciplinary boundaries (Arciniegas and Wood, 2006). Formally, GIScience has not defined a comprehensive list of journals that can be considered as common forums for publishing related work. Nevertheless, Fisher (2006) offers a non-exhaustive list of journals related to International Journal of Geographical Information Science (IJGIS), a well-known core journal for the discipline. Among them, *Photogrammetry Engineering and Remote Sensing*, *Geomatica*, *Remote Sensing of Environment*, *Geographical Analysis*, *Environment and Planning A*, *Environment and Planning B*, *Environment and Planning C*, *Computers & Geosciences*, *International Journal of Remote Sensing and the Photogrammetry Record*. The list is made up of a combination of GIScience recognized journals such as *Geomatica* or *Computers & Geosciences* and journals from allied disciplines such as *Environment and Planning A, B and C* or *Remote Sensing* journals. The list shows how GIScience is an intrinsically multidisciplinary field that exhibits a community of scholars who undertake work with a cross-disciplinary coverage and then publish their findings in a wide range of journals. However, as Fisher (2006) mentions, the list is neither a complete nor an exhaustive compilation of journals in the discipline. Thus, this study presents a new method for delineating the discipline in terms of journals tailored to GIScience's research nature.

3.2.3 Expanding the collaboration networks in GIScience

As the overall aim of this study is to represent and understand GIScience through its collaboration networks, first one needs to identify the set of journals that best represents the discipline.

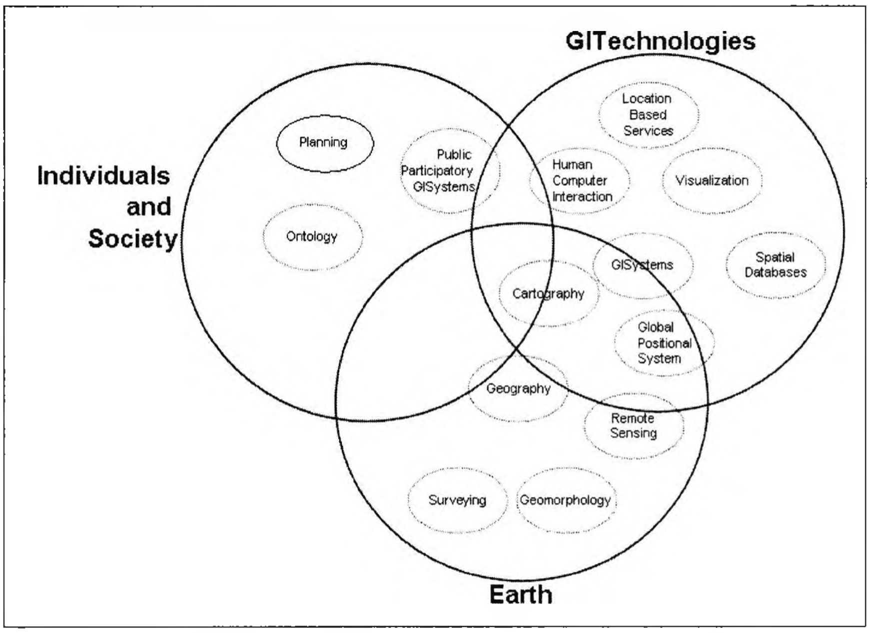


Figure 3-4 GIScience Domain Representation (based on Cova, 2000)

Cova (2000) and Goodchild *et al* (1999) explain that GIScience comprises a threefold interplay between society, GITechnologies and the Earth. As a result, a number of sub-specialities may emerge (see Figure 3-4), in accordance with the three components' inter-relationships. Hence, one cannot be certain about which may be most the representative publishing outlet, as the discipline not only comprises basic research (core journals that are easier to identify), but also to a certain extent, applied research in many other fields where GIScience has proved to be relevant. Thus, the list of journals, in which applied GIScience research findings and results may be published, cannot be clearly asserted. Therefore, using all journals classified under any of the fixed categories used in the JCR Social Science database from the Web of Knowledge (WoK) by the ISI is not convenient. Back in 2001, Fisher (2001) reported within ISI's categories 30 journals in the Geography, which has 38 today, and 54 journals under Information Science and Library Science that contains only one journal more nowadays. Therefore, none of the established categories with the ISI databases such as Geography (with 38 journals), Information Science & Library Science (55), Environmental Studies (51), Planning and Development (38), Social Sciences Interdisciplinary (59), Urban Studies (28) or Transportation (11) satisfactorily fits GIScience's scope and characteristics. On the other hand, if one would subjectively decide to take one, two or more related categories to represent GIScience, as they represent broader disciplines than GIScience, it may be possible that not all journals are related to the discipline. Thus, one needs to find a less subjective and selective mechanism to build GIScience's bibliographic data sources. The following sections elaborate on the process of assembling GIScience's publishing outlets.

3.2.3.1 Selecting a Core Journal

One assumes that scholars that publish in a core GIScience journal may carry out either applied or basic research; the two research streams observed by GIScience. Thus, core GIScience journals can be considered the common publication forum between the specialties surrounding the discipline, and authors publishing there are part of the collaboration network that may be built from this scientific literature. Therefore, the proposed approach starts with selecting a core journal that not only acts as a universal forum for scholars and researchers within the discipline, but also, ensures universality of the topics. The *International Journal of Geographic Science* (IJGIS) was selected because:

IJGIS holds the name of the discipline and is widely recognized as the primary academic journal for those working on GIScience issues (Fisher, 1996). It started in 1986, when only *Computers, Environment & Urban Systems* and *Geo-processing* were publishing on GISystems (Fisher, 2006). The former is more a computer-oriented journal, which now publishes a higher percentage of GIScience papers (Fisher, 2006), and the latter stopped publishing in 1985. Thus, IJGIS has been focused on GIScience topics during the longer time period.

3.2.3.2 Expanding the GIScience Publication Outlet

Authors publishing in IJGIS are only a part of the research collaboration in GIScience as authors may publish in other more topic-specific journals appropriate to their particular research areas. Thus, to expand the network the approach starts identifying *well-established authors* - defined as the most productive scholars in terms of numbers of published papers in the core journal, in this specific case, IJGIS. One argues that well-established authors are more likely to have a consistent tendency to work and then, publish papers in GIScience than authors with fewer papers. After identifying well-established authors, the next step is to determine the commonest set of publishing forums among them. As this type of author can be taken as working on relevant issues within the discipline, the commonest publishing forums will capture a wide range of core or applied topics that one assumes are strongly related to the discipline. There, one will be able to identify another core of allied field journals, where GIScience's principles or methods may be influential or useful.

To expand the GIScience's data source, two variables are used, *authors' productivity (ap)* and *journal relevance (jr)*. The former is calculated as a simple raw frequency that counts the total number of papers per author in IJGIS during a specific time period. After, setting up a convenient threshold for this variable, a list of well-established authors is assembled. A second variable, *journal relevance* counts the number of times a pair of well-established authors has published in a specific journal (it does not imply joint works). Finally, a threshold is also set up for this variable to capture the commonest publishing instead of occasional forums within the final list of journals, from which the final co-authorship network in GIScience will be assembled.

Alfred Lotka hypothesized that the number of authors publishing a given number of articles may correspond to a ratio with the number of authors publishing a single article (Lokta, 1926),

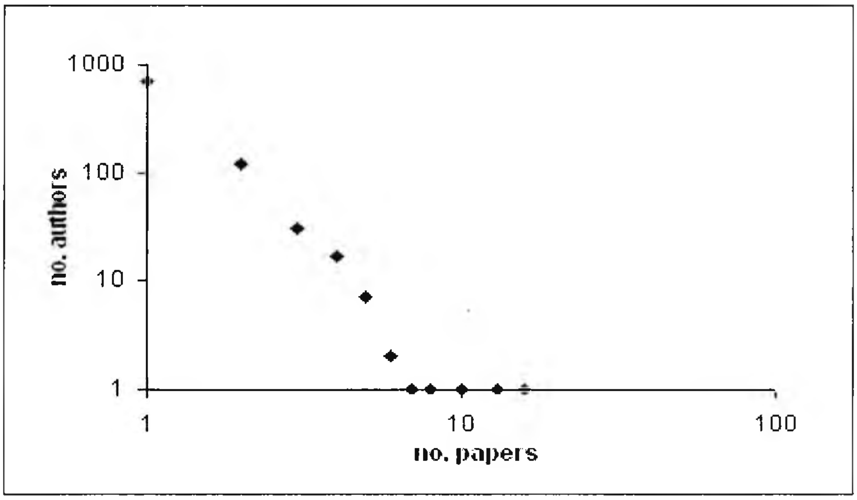


Figure 3-5 Log-Log plot showing the IJGIS paper distribution between 1992-2002

Thus, according to Lotka, as more papers are published, authors contributing with that many publications become less frequent. Consequently, the frequency of publication is not the same for all authors. During 1992-2002 (the time period of the study), 892 authors published in total 626 papers in IJGIS. In accordance with Lotka's Law, the paper distribution in Figure 3-5 reveals that only around 3% of the authors contributed more than 4 papers. In contrast, around 80% published only one paper.

The presence of a power law distribution in Figure 3-5 corroborates the presence of a small group of *well-established* author that show a consistent tendency to work and then, publish papers in GIScience during the time period. In contrast, the large group of one-contribution authors may correspond to either one-off GIScience involvement or new GIScience researchers. However, the former may correspond to new comers by a smaller degree as the study covers a long time period. Based on the paper frequency distribution values (see Figure 3-5), 4 and 5 were set up as candidate for the minimum number of publications required to be considered as a well-established authors in IJGIS.

No. Publications	Author Name	No. Publications	Author Name
16	goodchild, mf	4	buttenfield, b
13	burrough, pa	4	coppock, t
12	egenhofer, mj	4	defloriani, l
8	aangeenbrug, rt	4	florinsky, iv
8	fisher, p	4	griffith, da
7	martin, d	4	haining, r
7	stuart, n	4	hunter, gj
6	frank, au	4	kainz, w
6	heuvelink, gbm	4	kraak, mj
5	carver, sj	4	lees, b
5	govers, g	4	leung, y
5	jankowski, p	4	lovett, aa
5	jones, cb	4	masser, i
5	lee, j	4	medyckyjscott, d
5	mark, dm	4	okabe, a
5	mather, pm	4	peuquet, dj
5	muller, jc	4	puppo, e
5	shi, wz	4	raper, j
5	wise, s	4	sadahiro, y
5	worboys, mf	4	skidmore, ak
4	abel, dj	4	unwin, d
4	batty, m	4	wilkinson, g
4	brunsdon, cf	4	williamson, ip
		4	yeh, ago

Table 3.3 List of authors with more than 4 papers in IJGIS (between 1991-2003[1])

Table 3.3 shows the list of authors that have contributed with at least 4 papers in IJGIS. The list includes names such as Egenhofer, Heuvelink, Burrough, Goodchild, Worboys, Carver and Skidmore, some of whose works were the 10 most acknowledged (cited) works in IJGIS during 1987-2001 (Fisher, 2001). Assembling the list, a problem

[1]The initial calculations of author productivity included the years 1992 and 2003. However, the time period was shortened to 1992-2002 as problems with bibliographic data for these two years were found.

concerning the lack of standardization of author names was identified. A lack of author name standardization makes it difficult to fully identify an author because different sets of *surname* and *initials* may be used in different papers. In this case, it caused some papers to be ascribed to the wrong author, lowering the given author's total. At this stage, the solution was a manual time consuming double checking process for ambiguous authors' names that included browsing authors' personal web pages and comparing the affiliation addresses of the papers to be certain than repeated sets of surname and initials correspond to the same person.

The next step was to extend GIScience publication outlets by finding the common journals in which to the most prolific authors in IJGIS also published. A query yielded a total of 173 journals (see Appendix A) where authors in Table 3.3 have published at least one paper. The frequency distribution of the author participation (in Table 3.4) within these journals shows that while one journal had 18 contributions from different authors, 101 journals had only one top IJGIS author, suggesting one-off instead of regular contributions.

Authors Participation (no. papers published in)	**No Journals**
18	1
13	2
10	2
9	1
8	2
7	1
6	5
5	6
4	7
3	12
2	33
1	101

Table 3.4 Frequency distribution of the number of top-authors participation in other journals rather than IJGIS from journals indexed in ISI-WoK between 1991-2003.

Thus, one assumes that these 101 journals do not represent central research outlets for this group of authors, and are not much related to GIScience as the time period of more than 10 years is long enough to capture any representative publication trend. To assure commonality around topics covered by the GIScience publishing forum, a threshold of at least 2 different top-authors (*journal relevance* >= *2)* were set for selecting the journals, and as a result 72 journals were chosen(see Table 3.5).

Journal Name	No. Top-IJGIS Authors	Journal Name	No. Top -IJGIS Authors
Environment and Planning B-Planning & Design	18	Spatial Information Theory: A Theoretical Basic for GIS. LNCS	3
Computers & Geosciences	13	Technometrics	3
Environment and Planning A	13	Visual Computer	3
International Journal of Remote Sensing	10	ACM Transactions On Graphics	2
Photogrammetric Engineering and Remote Sensing	10	Advances in Spatial and Temporal Databases. LNCS	2
Progress in Human Geography	9	Agriculture Ecosystems & Environment	2
Annals of the Association of American Geographers	8	Canadian Geographer-Geographer Canadien	2
Computers Environment and Urban Systems	8	Catena	2
Geographical Analysis	7	Cities	2
Applied Geography	6	Computational Statistics	2
Cartographic Journal	6	Computer-Aided Design	2
Geoinformatica	6	CVGIP-Graphical Models And Image Processing	2
Journal of Geography in Higher Education	6	Ecological Modelling	2
Spatial Information Theory. LNCS	6	Environmental Management	2
Advances in Spatial Databases. LNCS	5	Forestry	2
Area	5	Futures	2
Geographical Journal	5	Habitat International	2
Geography	5	Hydrological Processes	2
Journal of the Royal Statistical Society Series D-The Statistician	5	IEEE Multimedia	2
Transactions of the Institute of British Geographers	5	IEEETransactions on Knowledge And Data Engineering	2
Cartography and Geographic Information Systems	4	International Journal of Climatology	2
Computer Journal	4	Interoperating Geographic Information Systems. LNCS	2
Earth Surface Processes and Landforms	4	ISPRS Journal of Photogrammetry and Remote Sensing	2
Environment and Planning C-Government and Policy	4	Journal of Environmental Management	2
Fuzzy Sets and Systems	4	Journal of Epidemiology And Community Health	2
Regional Studies	4	Journal of Hazardous Materials	2
Urban Studies	4	Journal of Hydrology	2
Computers & Graphics	3	Journal of Information Science	2
Environmental And Ecological Statistics	3	Land Use Policy	2
Geoderma	3	Landscape And Urban Planning	2
Geomorphology	3	Papers in Regional Science	2
Journal of Public Health Medicine	3	Pattern Recognition Letters	2
Journal of Regional Science	3	Regional Science and Urban Economics	2
Journal of The American Planning Association	3	Remote Sensing of Environment	2
Journal Of Visual Languages and Computing	3	Risk Analysis	2
Professional Geographer	3	Transportation Research Part C-Emerging Technologies	2

Table 3.5 List of 72 journals indexed in ISI-WoK in which top-IJGIS authors published more than 2 papers between 1991-2003.

Table 3.5 reveals some disciplines and research areas widely associated with GIScience such as Computing, Environmental Sciences, Remote Sensing, Geosciences or Cartography. Some, but not all, listed journals are classified as related journals to the discipline by Fisher (2006). Additionally the list suggests a less common set of journals such as *Journal of Epidemiology and Community Health or Risk Analysis* with only two IJGIS authors' contributors each. The topics covered by the list are in accordance with the multidisciplinary nature of the field and the interrelations of the three-fold discipline's components Society, Earth and GITechnologies topics, elucidated in section 3.1.2.

3.3 Assembling GIScience Collaboration Networks

Not all the topics covered by the journals in Table 3.5 are related to GIScience to the same extent. One expects to find some journals with bigger audiences and a higher number of contributors from GIScience than others. Therefore, one may speculate that those journals which are closer to GIScience and IJGIS would reveal a more interconnected collaboration network than those less related. Thus, the aim is to build six different collaboration networks and analyze their closeness to GIScience.

3.3.1 The Six Multidisciplinary Journal Sets

In contrast to methods for delineating disciplines explained in section 3.2.1, this study proposes an approach that is neither based on citation nor on centrality measures. Instead, it employs co-authorship links coupled to the use of two variables as selectors of journal relatedness to GIScience. In doing so, the method, overcomes the rigid ISI categories and is selective in choosing journals that are likely to publish GIScience related work.

Correspondingly, the strength of the journal-GIScience relationship is determined by authors' productivity and journal relevance. Author productivity (*ap*) determines whether an author can be taken as *well-established* in the core journal. In doing so, *ap* considers authors who have participated with at least 2, 3, and 4 papers in IJGIS during the time period. Journal relevance (*jr*) determines a minimum number of IJGIS well-established authors that a journal has had as contributors. To ensure commonality around GIScience topics, a journal has to have at least 4 different authors to be considered as part of the data source.

Author Productivity (*AP*)	**Journal Relevance (*JR*)**	
	JR = 5	**JR = 4**
AP = 2	35	72
AP = 3	21	39
AP = 4	10	27

Likeness to work on topics related to GIScience

Figure 3-6 Number journals in each data set and their likeness to represent GIScience publication outlets according to the values of ap and jr. Each data set is named after the values taken for *(jr, ap)* pairs, the labels 4-2, 4-3, 4-4, 5-2, 5-3 and 5-4 will be used to refer to each data source. The total number of journals in each data source includes 3 Lecture Notes on Computer Science that were unified later under one journal, but does not include IJGIS.

Combining the two variables *(jr, ap)*, six different data sets were assembled labelled as (4,2), (4,3), (4,4), (5,2), (5,3) and (5,4). Each data set comprises different journals selected according to the values of the two variables. Thus, Figure 3-6 displays the number of journals selected for each combination according to *(jr,ap)* values. As shown in Figure 3-6, 4-2 is the largest data source with 72 journals in which at least 4 different well-established IJGIS authors have published a minimum of 2 papers. In the same way, 5-4 shows that there are only 10 journals in which more than 4 top IJGIS authors have published more than 5 papers.

As Figure 3-6 indicates, the likelihood that a group of journals represents research areas or fields relevant to GIScience increases from right to left and from top to bottom, as the values of both variables increase. The 10 journals in 5-4 are more likely to represent relevant GIScience publication outlets as they covered topics with higher commonality to the discipline than the 72 journals in 4-2. Table 3.6 lists the journals in each data source including IJGIS.

Journal Title	(4, 2)	(4,3)	(4,4)	(5, 2)	(5,3)	(5,4)
ACM Transactions on Graphics						
Agriculture Ecosystems & Environment						
Annals Of The Association of American Geographers						
Applied Geography						
Area						
Canadian Geographer						
Cartographic Journal						
Cartography and Geographic Information Systems						
Catena						
Cities						
Computational Statistics						
Computer Journal						
Computer-Aided Design						
Computers & Geosciences						
Computers & Graphics						
Computers Environment and Urban Systems						
CVGIP-Graphical Models and Image Processing						
Earth Surface Processes and Landforms						
Ecological Modelling						
Environment and Planning A						
Environment and Planning B-Planning & Design						
Environment and Planning C-Government and Policy						
Environmental and Ecological Statistics						
Environmental Management						
Forestry						
Futures						
Fuzzy Sets and Systems						
Geoderma						
Geographical Analysis						
Geographical Journal						
Geography						
Geoinformatica (34 papers) + Transaction in GIS (84)						
Geomorphology						
Habitat International						
Hydrological Processes						
IEEE Multimedia						
IEEE Transactions on Knowledge And Data Engineering						
International Journal of Climatology						
International Journal of Geographical Information Science						
International Journal of Remote Sensing						
ISPR Journal of Photogrammetry And Remote Sensing						
Journal of Environmental Management						
Journal of Epidemiology And Community Health						
Journal of Geography In Higher Education						
Journal of Hazardous Materials						
Journal of Hydrology						
Journal of Information Science						
Journal of Public Health Medicine						
Journal of Regional Science						
Journal of The American Planning Association						
Journal of The Royal Statistical Society Series D-The Statistician						
Journal of Visual Languages And Computing						
Land Use Policy						
Landscape and Urban Planning						
Lecture Notes in Computer Sciences (Lecture Notes in Computer Science data are unified from the journals Spatial Information Theory, Advances in Spatial Databases, Spatial Information Theory: A Theoretical Basic For GIS, Advances						

in Spatial And Temporal Databases Proceedings and Interoperating Geographic Information Systems)						
Papers in Regional Science						
Pattern Recognition Letters						
Photogrammetric Engineering and Remote Sensing						
Professional Geographer						
Progress in Human Geography						
Regional Science And Urban Economics						
Regional Studies						
Remote Sensing of Environment						
Risk Analysis						
Technometrics						
Transactions of The Institute of British Geographers						
Transportation Research Part C-Emerging Technologies						
Urban Studies						
Visual Computer						

Table 3.6 Assembled publication outlets used to delineate GIScience collaboration networks. Dark grey highlighted the presence of a journal in any of the 4-X series and light grey in the 5-X series. Note that the 11 journals in 5-4 are part of all data sources.

3.3.2 Building the GIScience Co-Authorship Networks

The bibliographic information corresponding to the 72 selected journals between 1992-2002 was drawn from ISI-WoK. Then, as one file the information was parsed by a Java tool specially designed to construct the author networks. This section describes the networks assembling and the problems encountered during this process.

Exp. Tag	Represent	Data Element
FN	File type	ISI Export Format
VR	File format version number	1
PT	Publication type (e.g., book, journal, book in series)	Journal
AU	Author(s)	De Bruin, S
		De Wit, AJW
		Van Oort, PAJ
		Gorte, BGH
TI	Article title	Using quadtree segmentation to support error modelling in
		categorical raster data
SO	Full source title	INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE
DT	Document type	Article
C1	Research addresses	Wageningen UR, Ctr Geoinformat, POB 47, NL-6700 AA Wageningen,
		Netherlands
		Wageningen UR, Ctr Geoinformat, NL-6700 AA Wageningen, Netherlands
		Delft Univ Technol, Fac Aerosp Engn, Delft, Netherlands
ID	Author keywords	CLASSIFICATION ACCURACY; UNCERTAINTY; SIMULATION; PROGRAM;
		IMAGERY; IMPACT; PIXEL
BP	Beginning page	151
EP	Ending page	168
PG	Page count	18
JI	ISO source title abbreviation	Int. J. Geogr. Inf. Sci.
PY	Publication year	2004
PD	Publication date	MAR
VL	Volume	18
IS	Issue	2
GA	Part number	760PA
RP	Publisher web address	De Bruin S
		Wageningen UR, Ctr Geoinformat, POB 47, NL-6700 AA Wageningen, Netherlands
J9	29-character source title abbreviation	INT J GEOGR INF SCI
UT	ISI unique article identifier	ISI:000187841700003
ER	End of record	

Table 3.7 Example of an exported record from ISI-WoK, corresponding to a paper published in IJGIS during 2004. The information can be saved as a TXT file.

3.3.2.1 Raw Bibliographic Data

Despite the ISI-WoK database does not index some journals, conference proceedings and relevant books in GIScience. It was selected as the principal data source due to its wide and comprehensive coverage and because it is one of the few sources that holds information about author affiliations (Bordons and Gomez, 2000), information needed to geo-reference the collaboration networks. Table 3.7 is an example of how a single record that represents a journal paper is exported as a text file from the ISI-WoK. An exported record from ISI-WoK , as seen in Table 3.7, comprises tags and their corresponding information. The useful information for this study concerns AU = authors, SO = journal where the article was published, DT = type of the document, C1 = researchers' addresses, PY = year of publication, RP = address of the first author and ER tag that denotes the end of the current record.

3.3.2.2 Assembling the Co-authorship networks

As each one of the 6 networks comprises a different number of journals (see totals Figure 3-6), their sizes in terms of number of papers vary too. Table 3.8 shows the total number of papers that constitutes each network. The following tables show the total number of papers downloaded per journal.

Journal Title	Acronym	Total No. Papers
ACM Transactions On Graphics	ACM_TOG	234
Agriculture Ecosystems & Environment	AEE	1297
Annals Of The Association Of American Geographers	AAAG	1183
Applied Geography	AG	463
Area	AREA	1094
Canadian Geographer-Geographer Canadien	CANG	660
Cartographic Journal	CJ	352
Cartography And Geographic Information Systems	CGIS	56
Catena	CATENA	672
Cities	CITIES	635
Computational Statistics	CS	186
Computer Journal	CMJ	715
Computer-Aided Design	CAD	915
Computers & Geosciences	CG	1223
Computers & Graphics	COM_GRA	562
Computers Environment And Urban Systems	CEUS	169
CVGIO-Graphical Models And Image Processing	GMIP	135
Earth Surface Processes And Landforms	ESPL	914
Ecological Modelling	EMOD	
Environment And Planning A	EPA	2180
Environment And Planning B-Planning & Design	EPB	1019
Environment And Planning C-Government And Policy	EPCGP	583
Environmental And Ecological Statistics	EES	186
Environmental Management	EN	1075
Forestry	FOR	394
Futures	FUTURES	151
Fuzzy Sets And Systems	FSS	3393
Geoderma	GEOMODERNA	1137
Geographical Analysis	GA	267
Geographical Journal	GJ	1198
Geography	GEOGRAPHY	1240
Geoinformatica + Transactions in GIS	GEOINFORMATICA	118
Geomorphology	GEOMORPHOLOGY	970
Habitat International	HABITAT	83
Hydrological Processes	HP	1435
IEEE Multimedia	IEEE_MULTIMEDIA	377
IEEE Transactions On Knowledge And Data Engineering	IEEE_TKDE	867
International Journal Of Climatology	IJCL	1014
International Journal Of Geographical Information Science	IJGIS	626
International Journal Of Remote Sensing	IJRS	3015
ISPR Journal of Photogrammetry And Remote Sensing	ISPRS_JPRS	363
Journal Of Environmental Management	JEM	764
Journal Of Epidemiology And Community Health	JECH	1725
Journal Of Geography In Higher Education	JGHE	477
Journal Of Hazardous Materials	JHM	1360
Journal Of Hydrology	JH	2655
Journal Of Information Science	JIS	571
Journal Of Public Health Medicine	JPHM	1222
Journal Of Regional Science	JRGS	781
Journal Of The American Planning Association	JAPA	1367
Journal Of The Royal Statistical Society Series D-The Statistician	JRSSS	218
Journal Of Visual Languages And Computing	JVLC	250
Land Use Policy	LUP	416
Landscape And Urban Planning	LUPLA	1974
Lecture Notes In Computer Sciences	LNCS	266
Papers In Regional Science	PIRS	252
Pattern Recognition Letters	PRL	1604
Photogrammetric Engineering And Remote Sensing	PERS	1355
Professional Geographer	PG	1291

Progress In Human Geography	PHG	1572
Regional Science And Urban Economics	RSUE	444
Regional Studies	RGS	1406
Remote Sensing Of Environment	RSE	1289
Risk Analysis	RA	1171
Technometrics	TECNO	476
Transactions Of The Institute Of British Geographers	TIBG	667
Transportation Research Part C-Emerging Technologies	TRPCET	188
Urban Studies	UR	1930
Visual Computer	VC	324
Total		61171

Table 3.8 Complete list of journals, their acronyms and total number of papers between 1992-2002.

Table 3.8 shows that both the publication frequencies and also the number of years covered by the ISI-WoK database vary from journal to journal. FSS, IJRS, JH and EPA are journals contributing to the networks with more than 2000 papers, however, each journal's presence is not constant in all data sources. EPA and IJRS are part of all networks, while FSS is in all but 5-4 and JH only in 4-2. Table 3.9 reveals the difference in size between all networks.

4-2	**4-3**	**4-4**
61171	36167	25671
5-2	**5-3**	**5-4**
30983	21352	12457

Table 3.9 Total number of papers in each network data source

It is important to note that the networks' data sources in each series (4-X and 5-X) are nested. From left to right, 4-2 and 5-2 are the largest networks that contain all papers in 4-3 and 5-3 respectively. Similarly, all papers in 4-4 and 5-4 networks are part of the 4-3 and 5-3 data sources. Thus, 4-4 and 5-4 networks are contained in all subsequent networks in their respective series.

3.3.2.3 Co-Authors Paring Process

The process of assembling a co-authorship network consists of pairing all co-authors of a paper, repeating the process for all papers in the network's data source. As the smallest network comprises almost 13,000 papers, doing the pairing process manually is not a realistic task. Thus, a Java program was created for reading the ISI-WoK raw bibliographic data to make the collaboration combinations.

The method *readDataSource* (displayed in Figure 3-7), part of the java Class *JournalDatabase*, was designed to read the raw data. As the data source was organized in an Excel file, the first step of the reading method is to establish an OBDC connection to the physical *xls* file. Then, reading each paper and selecting key variables such as title, authors, journal source, addresses and publication year. As shown in Figure 3-7, the reading process consists of nested *if-statements*, as the exported raw data from ISI-WoK information is not organized in columns but in rows. A record that represents a paper is finished when an ER tag is encountered. The extracted information of useful tags is stored in vector structures as they will be used later by the *author-name replacement process*.

```
public void readDataSource()
{
 try
 {
 Class.forName("sun.jdbc.odbc.JdbcOdbcDriver");
 dbConnection= DriverManager.getConnection( "jdbc:odbc:Test1" );
 sqlQuery = dbConnection.createStatement();
 //Select all data within all worksheet in the Excel file associated.
 for (int k=0;k<numberSourceSheets;k++)
 {
 ...
 String sqlQueryString = new String("Select * from [Sheet"+a+"$]");
 queryResult = sqlQuery.executeQuery(sqlQueryString);
 //Query's result represented as a table
 resultTable = queryResult.getMetaData();
 numberOfColumns = resultTable.getColumnCount();
 //Read the information under each tag
 String previousTAG="";
 while (queryResult.next())
 {
       //Index j manages the columns within the jorunalMatrix
       for (int j = 1; j <= numberOfColumns; j++)
       {
        String columnValue = queryResult.getString(j);
        //Add an element to the tagsVector
        if (j == 1)
        {
        if ((columnValue == null) && previousTAG.equals("AU"))
  {
        columnValue = "AU";
        }elseif ((columnValue == null) && previousTAG.equals("TI"))
        {
               columnValue = "TI";
        }
        elseif ((columnValue == null) && previousTAG.equals("C1"))
        {
               columnValue = "C1";
        }
        elseif ((columnValue == null) && previousTAG.equals("DE"))
              {
               columnValue = "DE";
               }
              elseif ((columnValue == null) && previousTAG.equals("ID"))
              {
               columnValue = "ID";
               }
               elseif (columnValue == null) && previousTAG.equals("RP"))
               {
               columnValue = "RP";
               }
               else if (columnValue == null)
               {
               columnValue = "Blank";
               }

               tagsVector.add (columnValue);
               previousTAG = columnValue;
        }
        } //For-End
   }//While-End
 }//End-For K
 sqlQuery.close(); dbConnection.close();
 }
 catch(Exception ex) { … }
}//End-Method
```

Figure 3-7 `readDataSource()` Java Method, part of the main Java application, for reading the ISI-WoK raw data.

After reading the data, the following step used all strings of authors extracted for each paper to create the co-authorship links that represent the scientific collaboration within the given paper. The essence of the method `makeAuthrosCollaborationsCombinations` is to make binomial combinations between all co-authors of the papers. The binomial coefficient (BC) is defined as

$$\text{BC of}\binom{n}{k} = \frac{n!}{k! * (n-k)!}, n \geq k \geq 0$$

Equation 3-2 Binomial Coefficient (BC)

In this specific case *n* is equal number of co-authors and *k* equals 2, representing the size of the combination. If a paper is written by 3 co-authors A, B, C, the total number of combinations is defined by

$$\text{BC of}\binom{3}{2} = \frac{3!}{3! * (3-2)!} = 3$$

The paper will be represented by 3 co-authorship combinations as follows: (A,B), (A,C) and (B,C). The same process is repeated for each paper and finally, equal pairs that represent the same collaboration are added to calculate the strength of that specific scientific collaboration pair. It is important to note that in the case of co-authorship pairs, (author A, author B) is equal to the (author B, author A) pair, which is not the case of citation pairs where (P1 cites P2) does not imply that (P2 cites P1). However before creating the final *(author1,author2) collaboration matrix*, the author names were parsed by a Java routine designed to deal with the lack of standardizing on author names. The following section presents the proposed solution.

3.3.3 Standardizing Author Names

Within the bibliographical context, an author is identified as a name set that comprises *surname, initials,* e.g. Walker, Ra or Fisher, PF. However, as authors have discussed (Horn *et al*, 2004; Moody, 2004; Newman, 2004; Elmacioglu and Lee, 2005 ; Murray, *et al*, 2006) and was detected during the selection of well-known IJGIS authors (see section 3.2.3), there is a problem due to the lack of standardization of authors naming. This *classical name authority control problem* (Elmacioglu and Lee, 2005) affects the validity of co-authorship, citation or co-citation analysis (Newman, 2004; Costas and Bordons, 2005; Calero *et al*, 2006) at author level, as it depends on the ability to fully identify authors, as their target analysis units. The problem arises when it is not possible to fully identify an author. The same author name may be written using various spellings such as Fisher, PF or Fisher, F (synonymous names). Equally, the same spelling may represent different authors (homonymous names); a common problem within authors from Oriental origins with very popular and short surnames such as Li, Chang or Shi.

3.3.3.1 The String Comparison Approach

Some of the solutions that have been proposed for the author names' consistency problem involve manually cleaning up small data sets (Murray *et al*, 2006); creating a list of all possible variants for all authors names applied to a list of 333 researchers from natural resources in Spain (Costas and Bordons, 2005), using a combination of authors initials and their main organization addresses to better identify authors within nanotechnology centres (Calero *et al*, 2006), or including the creation of two collaboration networks (Newman, 2004b), one that identifies authors with only one initial and the other that identifies authors using all available initials.

As the GIScience's bibliographic data set for this study comprises 60.000 papers with more than 66059 identified author names from the 72 selected journals, a manual check as carried out by Murray *et al* (2006) is impractical. So, this study proposes a new approach that uses full author names including all initials as Newman (2004b) and authors' affiliations as Calero *et al* (2006), but adding extra flexibility by measuring how similar authors' addresses may be.

In order to identify as fully as possible all authors in the data source, the measure is estimated using an assembled *text similarity function* based on the addresses under the RP and C1 tags (refer to Table 3.7 for more information on the tags) that implements the *Dice Coefficient (DC)* algorithm for matching string similarity (Cohen, *et al*, 2003). The DC returns a numeric value as the measure of similarity between two strings. To some extent, a DC value suggests that one match is better than another. The DC, according to Chapman (2005), can be defined as follow:

$$\text{Dice Coefficient (DC)} = \frac{(2 * CommonTerms)}{(NumberOfTermsString1 + NumberOfTermsString2)}$$

Equation 3-3 Dice Coefficient

The proposed solution uses *SimMetric* (Chapman, 2005), an open source Java library that calculates DC values between text strings that in this case represent authors' addresses. When comparing two authors' names with the same surname but different initials, a DC measure between addresses indicates how similar the two strings are. If DC equals 1 the strings are identical, and if DC is greater than a set threshold similar enough to be the same address. After all variants of the same author's name backed by his/her addresses, have been identified, a unique name is set up for unifying all variants corresponding to this name within the data sources.

3.3.3.2 The Author Name Replacement Java Tool

The proposed method described in the previous section was implemented in a Java class (see Figure 3-8 for the Class' highlights). First, it calculates the likelihood of two authors with the same surname and initials (or same surname and only one of the two initials) being the same individual based on how similar the two authors' addresses are. However, before comparing the two addresses, common or stopper words such as *centre, university, national, dept, etc.* are suppressed to stop getting high similarity between the two due to the presence of these words. The authors in the following example appear to be the same person as the two addresses are about 47% similar.

martin, d	univ southampton dept geog southampton so9 5nh england;	0.47
martin, r	univ cambridge dept geog downing cambridge cb2 3en england;	

Table 3.10 Presence of stopping words within the similarity measure

However, a simple check reveals that are in fact two different authors, and , in this case the similarity is biased by the presence of common words such as *univ, dept and geog.* Therefore, the proposed solution manages a list of stop words that are removed from the addresses before calculating the similarity measures. Nonetheless, two different addresses from authors within similar names do not directly imply two different authors. There may be the possibility that they are the same person, who has worked for different institutions or organizations over the years. To overcome this, the process relies on all authors' affiliation under C1 tag, as it was found that the first two addresses correspond to the first author from the RP tag.

```
public void calculatingSimilarity()
{
 for (int i = 0; i < tagRPVector.size()-1; i++)
 {
 // Get the primary key - author name.
 ...........
 // Composing the name divides into surname and initials
 ...........
 for (int j=i+1;j<tagRPVector.size();j++)
 {
 .....
 // Composing the name divide into surname and initials
 ........
 if (surName1.trim().equalsIgnoreCase(surName2.trim()))
 {
 //Extract address correspond to the second author.
 String originalAddress2=tokens2[1];;
 address2= suppresingWords(originalAddress2).trim();
 //Create the Similarity Record, Author1, Author2 and similarity measure
                    float                  similarity                   =
Compare.getSimilarity(address1.toLowerCase().trim(),address2.toLowerCase().tr
im());
 String name1 = surName1.trim()+", "+initials1;
 ......
 String firstInitial1=initials1.substring(0,1);
 //To know if the first initial of the first names is
 String firstInitial2=initials2.substring(0,1);
 //equal to the first initial of the second name
               if             ((similarity               !=0)&&(similarity
1.0)&&!(name1.equals(name2))&&(firstInitial1.equals(firstInitial2)))
  {
 //Create Similarity record to be added to the SimilarityVector Using
 //AU-AuthorName and FullAddresses.
  SimilarityRecord authorSimilarity = new SimilarityRecord(name1,address1,
originalAddress1, name2, address2,originalAddress2,similarity);
       ......
 }
       ...
 }//END-Method
```

Figure 3-8 `calculatingSimilarity` Java Method. It calculates a similarity measure (based on the Dice Coefficient) by comparing the addresses of authors who have equal *surname, initials sets* or have equal surname and one of the initials.

Once the similarity measure is calculated with the help of RP and C1 addresses, a replacement table is assembled with the ambiguous authors' names. The table lists similar authors from which the similarity measure between their addresses are higher than a set up threshold. In this case the threshold was set up as a minimum of 30% commonality (equal words) between two strings. Apparently, it appears to be a low threshold, but after experiments with higher threshold figures the similarity procedure did not identify many potential matches. The reason may be that similar to author names, there are no standards on naming organization addresses. Therefore, after suppressing stopping words from the C1 and RP, shortened forms of the addresses are left as very short strings. If the threshold is a greater value, only the exact strings will be identified, mismatched shorter strings that may represent the same places (see Bennett, r in the following table).

Each record in the replacement table (see Table 3.11) comprises an author name and a list of similar author names and their addresses.

Author goodchild, m	
	Author goodchild, mf Address univ calif santa barbara ncgia santa barbara ca 93106 usa
	Author goodchild, mf Address univ calif santa barbara santa barbara ca 93106 usa
	Author goodchild, mf Address univ calif santa barbara dept geog santa barbara ca 93106
	Author goodchild, mf Address univ calif santa barbara santa barbara ca 93106
	Author goodchild, mf Address univ calif santa barbara natl ctr geog informat & anal 3510 phelps hall santa barbara ca 93106

Author Bennett, r	
	Author bennett, rj Address london sch econ dept geog london england
	Author bennett, rj Address univ london london sch econ & polit sci ctr econ performance houghton st london wc2a 2ae England
	Author bennett, rj Address univ london london sch econ & polit sci dept geog london wc2a 2ae England
	Author bennett, rj Address univ london london sch econ & polit sci div geog london wc2a 2ae England
	Author bennett, rj Address univ london london sch econ & polit sci london wc2a 2ae England

Table 3.11 An example of two records in the replacement table.

For example, the addresses' records corresponding to M.F. Goodchild, listed in Table 3.11, are similar enough to the listed addresses of M. Goodchild. Thus, *Goodchild M.* will replace *Goodchild M.F.* in the corresponding records in the raw data file (see the structure of a data file raw record in Table 3.7). In that way, similar author names with similar enough addresses are unified under one name. In total, 6888 author pairs with the same surname and first initial but with different second initial were identified. From this total and according to the similarity calculations, only 3174 author pairs consisted of the same author. Thus, 2966 author names were replaced by 1013 author names (see examples in Table 3.11).

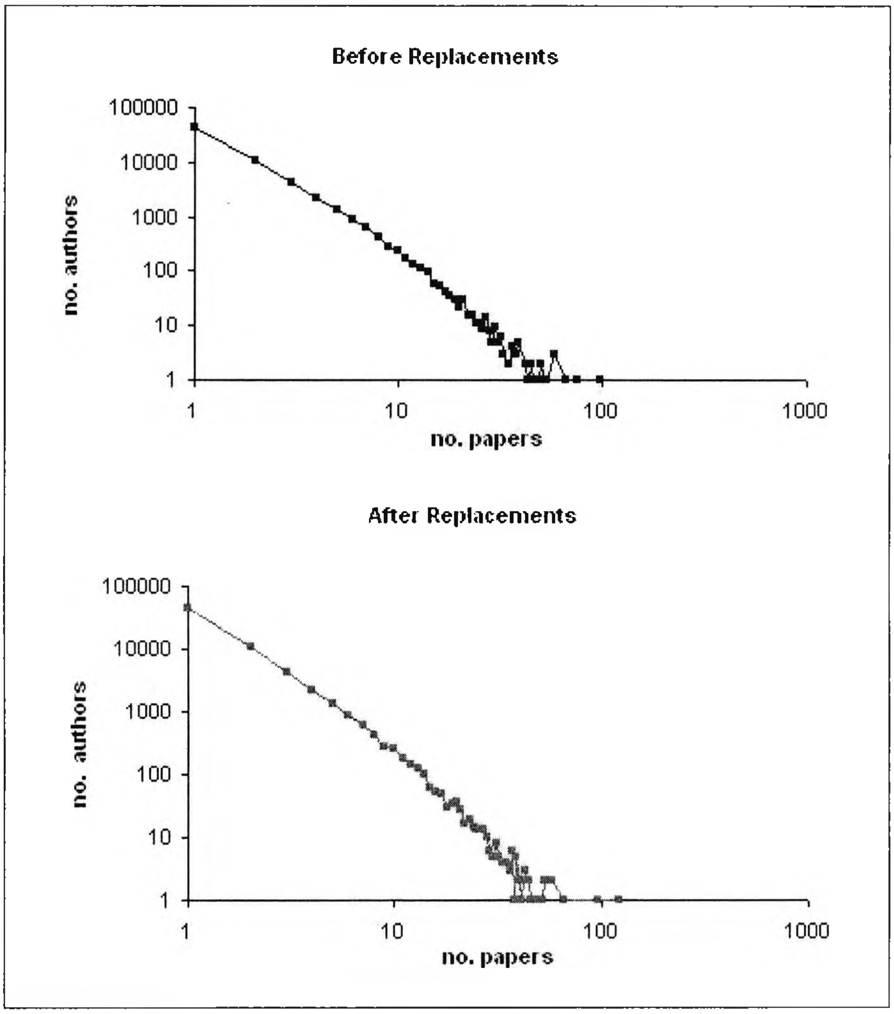


Figure 3-9 Log-Log Plot distribution of the number of papers per author corresponding to 4-2 data source. The top panel shows the distribution before replacements (in blue) and the lower panel after replacements (in pink).

The frequency distribution of the number of papers before and after replacements in Figure 3-9 reveals that after similar author names replacements (see section 3.3.3.2) there are 340 authors less with 1 papers, 189 authors less with 2 and 77 with 3 papers. It appears that this paper counts redistributed around authors with 10 to 25 papers. Moreover, comparing the network properties before and after the replacements their values vary slightly. The mean paper per author increases around 1% for all networks except for the 5-3, for which the increment is around 2% (from 1.91 to 1.94). The average distance between authors in the largest component decreases in all networks. Distances vary from 8.87 to 8.75 steps (1.4%) in network 4-2 to 11.02 to 10.44 steps (5.2%) for network 4-4. Thus, one can conclude that the GIScience networks representations are not sensitive to the problem poses by the lack of standardization of the author names.

3.3.3.3 Issues within the Replacement Tool

The information on authors' affiliation by paper from ISI-WoK appears to be a very useful piece of information within the process of fully identified author names. However, the fact that for a paper one cannot match one-to-one the C1 addresses to the authors' list makes the information less useful. The only way to match authors with address is using the RP address that corresponds to the first author within the paper. Therefore, the main drawback of the proposed tool is its reliance upon the RP field, because, if an author in the bibliographic data source has never been a first author, then he/she will not have addresses to compare to.

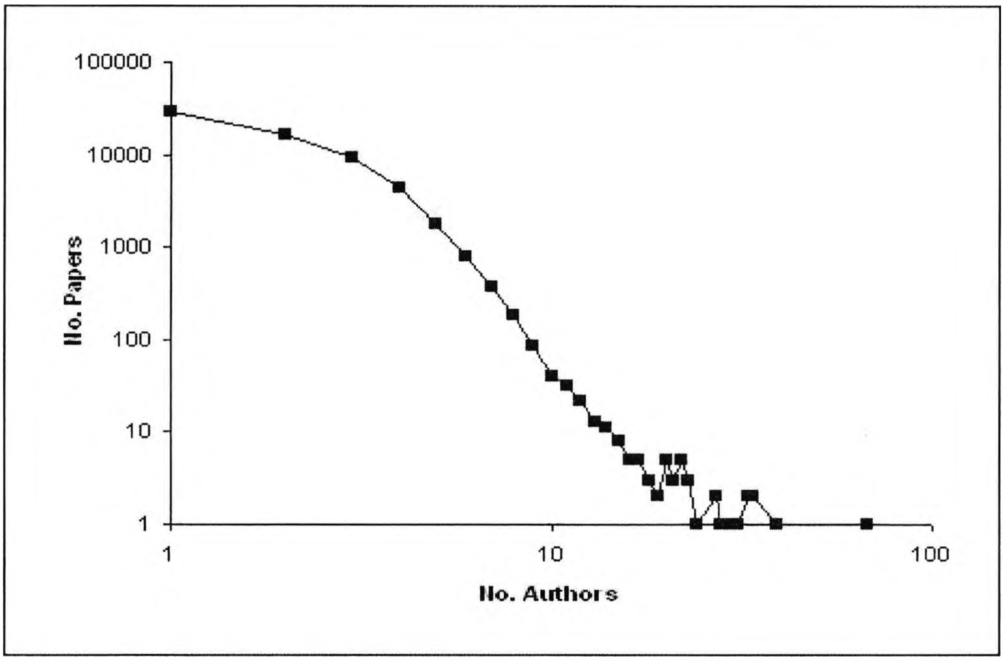


Figure 3-10 Log-Log Plot of the distribution of the number of authors in 4-2 data source.

Nevertheless, according to the distribution of number of authors per paper in Figure 3-10, around 46% of the papers within 4-2 data source (which contains all other data sources) are sole works. In this way, around 29,000 authors have been listed at least once as first author. Additionally, there are 1,600 authors with more than one paper that have been listed as first author. Thus, the corresponding RP and C1 addresses from around 50% out of the 66,000 authors in the data source can be used to clarify synonymous author names, if there would be the case.

Contrary, author cases with common surnames, especially those from oriental countries such as China, Japan or Korea, pose the main challenge. In this case, C1 addresses cannot help much as the list of possible academic organizations from these countries is not very long, accentuating the possibility that homonymous authors work at the same organization, making it even more difficult to fully identify them. However, the entries from papers having authors affiliated to these countries as co-authors represents around 8% of the total number of authors, which is not statistically significant if compared with the total number of authors.

This chapter delivers a new approach to build co-authorship networks from a highly multidisciplinary field such as a GIScience. The proposed approach does not rely on fixed subject categories offered from the majority of the bibliographic databases, but on choosing flexible data sources more in agreement with the dynamic characteristic of the field. If a new sub-field emerges from the field, authors in the domain will start making themselves visible by publishing on related topics in the discipline's core journals. Hence, the methodology will embrace the common new sources from well-established authors. The proposed approach builds six bibliographic data sources that represent different instances of scientific collaboration in GIScience. They range from a collection of 72 journals covering a wide range of core and applied topics to GIScience, to 11 well known journals for the field's research community. Additionally, traditional problems due to the lack of standardization in author naming were encountered. To overcome this, a method for identifying authors using their full set of initials and their addresses was designed and applied. However, restrictions on matching authors by their respective addresses hinder the scope of the solution. Nevertheless, as a result of using authors' names coupled with a similarity measure, the proposed method was able to identify around 7000 potential cases of ambiguous author name pairs. From them, around 3000 pairs were identified as the same author and replaced under a unique author name.

4 The Structure of Collaboration Networks in GIScience

This chapter explores the structure and the topology of each one of the six GIScience co-authorship networks built and described in the chapter three. The aim is to characterize each co-authorship network and to explore to what extent the topology of each network is determined and influenced by which topics are covered and also, by which are not covered. As built from different data sources, each network represents a unique insight into how co-authorship relationships connect scholars that publish work on topics ranging from fairly to highly relevant to GIScience. Therefore, one expects that each network will yield distinctive patterns of collaboration and publishing, and will convey different sets of central and influential authors within the academic society surrounding the covered research areas.

There are two essential outlines for the chapter's delivery and discussions. One is a summary of the basic properties of each network and second, is a comparison table of GIScience and the properties of other disciplines. Table 4.12 shows a summary of the main statistical properties.

	Collaboration Network					
Property	**4-2**	**4-3**	**4-4**	**5-2**	**5-3**	**5-4**
No. journals in the network data source	72	39	27	35	21	10
No. of authors' names Identified	65410	32348	22154	31678	19513	13950
No. papers written by the authors	62155	36167	25671	30983	21352	12457
Ratio authors/ papers	1.05	0.89	0.86	1.02	0.91	1.12
Mean authors per paper (co-authorship)	2.05	1.76	1.73	1.91	1.77	1.96
Mean papers per author (network participation)	1.95	1.97	2.0	1.87	1.94	1.75
% Authors with only one paper	67%	67%	68%	69%	68%	70%
% Sole papers (individual works)	46%	58%	58%	52%	57%	51%

Table 4.12 Summary of properties of authors and papers for six different GIScience co-authorship networks, between 1992 and 2002.

Table 4.13 shows the statistical properties of only two of the six GIScience networks together with results of others disciplines' collaboration networks. The 4-2 and 5-4 networks were chosen to be compared to similar studies done on other disciplines because they depict authors' scientific interrelations from opposite extremes regarding topics that cover GIScience (see Section 3.2.3 Chapter 3). The 4-2 network embraces a wide range of topics that may or may not be related to GIScience, while the 5-4 network portrays the interrelations of authors working on topics very relevant to GIScience.

	5-4 GIScCN	**4-2 GIScCN**	**Computer Science (preprints**	**Human Computer Interaction (HCI)**	**Bio-medical Research**	**Physics (online preprint database**	**Mathematics**	**Social Science**
Source	ISI-WoK	IS-WoK	NCSTRL	ACM	Medline	Physics E-Print Archive	Mathematics Review Journal	Sociological Abstracts
Number Authors	13950	65410	11994	23624	1520251	52909	253339	128,151
Number Papers	12457	62155	13169	22887	2163923	98502	-	281090
Mean Papers Per Author	1.7	1.9	2.6	2.2	6.4	5.1	6.9	na
Mean Authors Per Paper	2.0	2.1	2.2	2.3	3.8	2.5	1.5	2.4
Mean Number Collaborators Per Author	4.8	5.5	3.6	3.7	18.1	9.7	3.9	na
Size of Largest Component	31%	49%	57%	51%	92%	85%	82%	53%
Mean Distance	10.0	9.0	9.7	6.8	4.6	5.9	7.6	9.8
Largest Distance	28	26	31	27	24	20	27	na
Time Window	1992-02	1992-02	1995-99	1998-02	1995-99	1995-99	1995-99	1969-99
Work Ref.			(Newman, 2001a).	(Horn, et al., 2004)	(Newman, 2004c).	(Newman, 2004c	(Newman, 2004c	(Moody, 2004)

Table 4.13 Summary of the basic results of the analysis of eight co-authorship networks, 2 from GIScience and 5 from other fields, sub-fields and disciplines.

The disciplines compared vary from traditional sciences such as Mathematics and Physics, through relatively new fields or subfields like GIScience, such as Computer Science and Human Computer Interaction (HCI), to areas with very particular patterns of collaboration such as Biomedical research and Social Science. The statistical properties in the summary tables are explained in detail and cited throughout the content of the chapter.

This chapter is organized as follows. The first section focuses on studying the networks' basic statistics and topological properties in order to highlight any global collaboration and publishing patterns that help to characterize the group of scholars behind each network. The second and the third sections analyse to what extent co-authorship relations connect and place authors within the overall network structure.

Finally, the fourth section focuses on determining the degree to which each network structure is dependent on a group of central authors.

4.1 Basic Results and Statistics

This section aims to analyze the basic properties of the six academic networks built from a large body of literature written by scholars working on topics with varying relatedness to GIScience. In doing so, one expects to make evident any non-trivial *publication, collaboration and co-authorship trends* around research in GIScience that help us to understand the nature of its multidisciplinary scientific practice.

As this research has not placed any boundaries around a specific GIScience research agenda, the analysis of six distinct collaboration networks are windows into six different social academic structures. This study argues that highlighting any apparent differences within the basic properties of the network under study will help us to identify how similar/ dissimilar the academic network structures are. Moreover, it may also help us to clarify the complicated and vaguely defined (Goodchild, 1992; Goodchild, 2006) network of academic interrelations around GIScience, consequence of a highly multidisciplinary GIScience research agenda.

4.1.1 Papers and Authors' Statistics

The size of the networks, in terms of number of papers and authors, varies considerably depending on the journals covered by each data source. However, there is no evident trend other than the size of the networks depends on the number of journals included. Figure 4-11 makes evident the difference in the author-paper ratios for all networks. The 4-2 network built from 72 journals comprises almost five times more papers than the 5-4 network assembled from 10 journals. However, it is important to bear in mind that the number and nature of the journals included is the result of the process devised to build each GIScience co-authorship network data source (see section 3.3 chapter 3 for more detail). This process implies nested data sources, where larger networks embrace the smaller ones. On the one hand, 4-2 contains data from all other networks. In contrast, 5-4's data is part of all networks. Thus, only a small fraction of the papers is unique to each network.

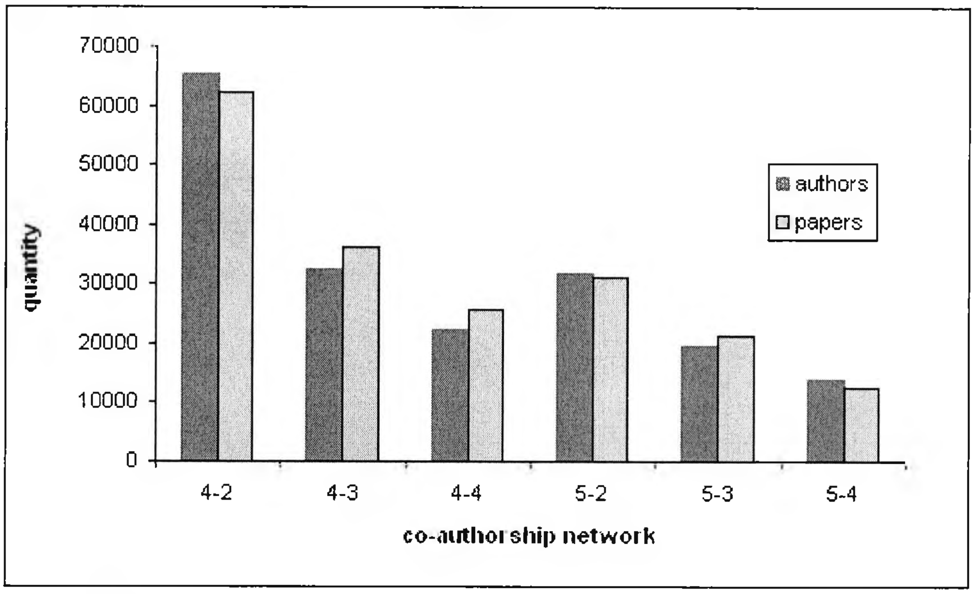


Figure 4-11 Histogram comparing the number of papers and authors for each co-authorship network.

The case of the total number of authors is much more complicated. As was mentioned in section 3.3.3, lack of standardisation in author naming makes the task of fully identifying authors very difficult (Newman, 2001a; Horn *et al*, 2004; Barabasi *et al*, 2004; Moody, 2004), and so, to calculate exactly their total number. Therefore, only the total number of distinct authors' names in a given data source can be identified. The problem, addressed in Chapter 3, generates mistaken author identification, where an author name may be shared by two different people, or an author may be represented by two different names. The results have shown it to be insignificant by comparing networks before-after corrections (see Chapter 3). Hence, for the context of this research total author names is assume to represent the total number of unique authors.

The total number of authors was calculated after applying the replacement method (see Chapter 3 - section 3.3.3.1). The results (see Table 4.12 second row) show that there is no a clear trend within the number of authors in all networks. There are networks such as 5-2 or 4-2, the largest in size, where the number of authors is larger than the number of papers. One possible explanation that may account for it is that papers covered by these networks are written by new authors in larger teams, rather than by those authors who were already in the network. Conversely, networks such as 4-3, 4-4 and 5-3 where there are more papers than authors, may depict collaborations from most universal topics for many, or depict less collaborative authors writing the papers. These premises are investigated in section 4.1.3., where the co-authorship and collaboration trends are analyzed.

4.1.2 Network Participation (mean published output)

Many studies carried out on scientific collaboration networks (e.g. Newman, 2002; Börner *et al*, 2004; Moody, 2004; Wagner and Leydesdorff, 2005) have used the *mean number of papers per author* to typify publication trends within the disciplines or fields under study. However in this specific case, the paper networks built are partial presentations of academic relations in GIScience as many relevant bibliographic sources and their co-authorship relations were not covered (chapter 3 Section 3.3.2). Evidence of this may be found out by considering publication lists from the authors with the major contributors to IJGIS. From the personal web pages from authors such as M. Goodchild, P.A. Burrough, M.J Egenhofer, P. Fisher, D. Martin, N Stuart, A.U. Frank or G.B.M. Heuvelink, it is evident the presence of a wide variety of non-covered journals such as Transactions in GIS, Journal of Geographical Systems, URISA or proceedings from well-established GIScience conferences. As a consequence, it is very likely that the networks' data sources do not represent the full list of any author' publications. In fact, in all networks more than 60% of the authors (see Table 4.12) contribute only with one paper, as can be seen on the left hand side of the authors' participation distributions in Figure 4-12.

Therefore, within the context of this research, the *mean papers per author* is regarded as a "*network participation*" rather than an author's productivity measure. Other studies have found that relatively new disciplines such as computer science and the HCI show similar low averages as the GIScience co-authorship networks studied here. Newman (2001a) investigating Computer Science, and Horn *et al*, (2004) reviewing HCI concluded that in their studies the low participation averages of 2.6 and 2.2 papers respectively can be a consequence of the small size (4 years) of the bibliographic repositories that is translated into a poor network coverage. However, despite covering publications for an 11 year time window, author participation in the GIScience collaboration networks only ranges from 1.75 to 2.0 papers. One can conclude that co-authorship GIScience networks, may be suffering from the same bias of coverage and database constraints as HCI and Computer Science networks. Conversely, results from co-authorship studies on disciplines such as Physics, Biomedicine or Mathematics (see Table 4.13) show much larger network participation, with averages ranging from 5.1, 6.4 to 6.9 papers per author. In contrast to GIScience, these well-defined disciplines have very comprehensive bibliographic databases that work as universal repositories with ample topic coverage. Thus, it is very likely that these databases have indexed a high percentage of the papers that have been published on these areas.

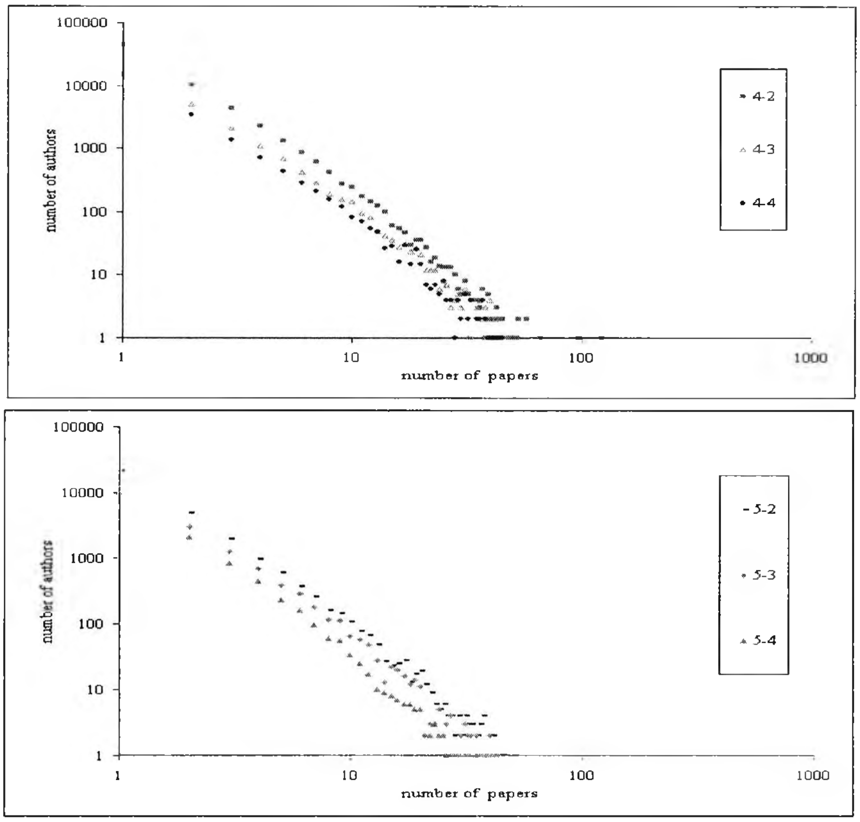


Figure 4-12 Author-Participation Histogram (log-log plot) from 1992 to 2002. Top panel shows the distribution for all 4-X networks. Bottom Panel shows the distributions for all 5-X networks

In addition to the large group of authors with one paper, author network participation distributions in Figure 4-12 show the presence of a small group of highly participative authors (tail's right end). From the 20 most participative authors' list in Table 4.14 some patterns can be highlighted. On one hand, there are highly participative authors in network 4-2 such as K. Beven, RK, White or J. Poesen, but who are not present at the top ranking of any other network. As the 4-3, 4-4 and any 5-X networks do not cover these journals, then, the participation of these authors decreases. They may represent the group of the *less likely GIScience authors* who publish in the least common journals such as Ecological Modelling, Habitat International or Journal of Epidemiology and Community Health.

	GIScience NETWORKS											
	4-2		4-3		4-4		5-2		5-3		5-4	
	Author	No. Papers	Author	No. Papers	Author	No. Papers	Author	No. Papers	Author	No. Papers	Author	No. Papers
1	Johnston, R	123	Johnston, R	122	Johnston, R	116	Johnston, R	120	Johnston, R	91	Johnston, R	67
2	White, Rk	97	Batty, M	57	Butler, J	53	Butler, J	54	Butler, J	53	Butler, J	53
3	Singh, Vp	66	Butler, J	53	Cracknell, Ap	52	Cracknell, Ap	52	Cracknell, Ap	52	Cracknell, Ap	52
4	Batty, M	58	Cracknell, Ap	52	Batty, M	48	Singh, Vp	51	Batty, M	45	Batty, M	44
5	Poesen, J	58	Buckley, Jj	45	Buckley, Jj	45	Batty, M	45	Buckley, Jj	45	Foody, G	32
6	Butler, J	53	Pattie, C	44	Pedrycz, W	44	Buckley, Jj	45	Pedrycz, W	44	Varotsos, C	31
7	Pedrycz, W	53	Pedrycz, W	44	Pattie, C	42	Foody, Gm	44	Wu, Cx	39	Longley, P	30
8	Cracknell, Ap	52	Taylor, P	42	Green, Dr	39	Pedrycz, W	44	Green, Dr	38	Curran, P	29
9	Jorgensen, Se	50	Clout, H	39	Wu, Cx	39	Taylor, Pj	40	Fuller, R	35	Karnieli, A	26
10	Curran, P	47	Green, Dr	39	Clout, H	37	Green, Dr	39	Foody, G	34	Atkinson, P	25
11	Taylor, P	46	Imrie, R	39	Hudson, R	37	Wu. Cx	39	Longley, P	34	Goodchild, M	25
12	Buckley, Jj	45	Wu, Cx	39	Longley, P	37	Clout, H	36	Goodchild, M	33	Clark, Gl	24
13	Oppenheim, C	45	Dorling, D	38	Taylor, P	37	Goodchild, M	35	Kandel, A	33	Fisher, P	24
14	Pattie, C	44	Martin, D	38	Fuller, R	36	Clark, Gl	33	Clark, Gl	32	Egenhofer, Mj	23
15	Smith, Gd	44	Clark, Gl	37	Ilbery, B	36	Kandel, A	33	Fisher, P	32	Gong, P	23
16	Beven, K	43	Hudson, R	37	Foody, G	35	Smith, Dm	32	Curran, P	31	Staab, Ca	23
17	Kirby, A	43	Longley, P	37	Martin, D	35	Castree, N	31	Mesiar, R	31	Castree, N	22
18	Walling, D	43	Fuller, R	36	Clark, Gl	34	Dorling, D	31	Varotsos, C	31	Dorling, D	22
19	Dorling, D	42	Ilbery, B	36	Curran, P	34	Hudson, R	31	Hong, Dh	30	Martin, D	21
20	Nijkamp, P	42	Curran, P	35	Dorling, D	34	Martin, D	31	Taylor, P	30	Fotheringham, As	20
21			Fisher, P	35	Fisher, P	34	Mesiar, R	31			Jensen, Jr	20
22			Foody, G	35							Pattie, C	20
23			Goodchild, M	35							Timmermans, H	20
24											Townshend, Jrg	20

Table 4.14 The 20 most participative authors (there may be more than 20 if the scores are equal). Light shaded authors exhibit a constant participation throughout all networks. Ron Johnston (dark grey shade) is the most participative author within all networks..

The opposite case shows a group of highly participative authors in the 5-4 network such as P. Atkinson, Mj, Egenhofer, As. Fotheringham, P. Gong, Jr. Jensen, A. Karnieli, JRG. Townshend or H. Timmermans, from whom works are mainly on core GIScience topics. Therefore, they form the group of *most likely GIScience authors*. Among them, the most distinctive sub-group comprises authors, highlighted in light grey, who are in all networks' rankings.

Accordingly, the two panels in Figure 4-12 show how the addition of new journals to each network increases the number of non-participative authors (see close to the y-axis), but also how the small group of highly participative authors, in the right-end tail, becomes even more participative in each subsequent network. Apart from a fat-tail consisting of small number of authors who publish a very large number of papers, the appearance of a straight line in a log-log plot is observed in Figure 4-12. The latter is taken as strong evidence that the plotted distribution follows a power law (Newman, 2005; Adamic, and Huberman, 2002). According to a power law, the size of x^{th} largest event scales with some property of the event in the form of

$$P(c) = C * x^{-\sigma}$$

Equation 4-4 Power Law

Where *C* is a constant, in most cases uninteresting (Newman, 2005), and σ denotes the exponent of the power. This exponent indicates how the distribution changes as a function of the underlying variable (*x*). Scholars have drawn attention to calculating the distribution's exponent. Many have taken it as the slope of a fitted line in a log-log plot. However, some authors (Adamic. and Huberman, 2002; Archambault , 2000; Newman, 2005) have argued that there is a tendency of this method to overestimate the slope of the power law (Newman, 2000). Moreover, as Figure 4-12 shows, sometimes distributions exhibit very noisy tails due to the low number of authors with large number of papers. Newman, 2005 proposes an alternative simple maximum likelihood method to calculate the exponent using the following formula:

$$\sigma = 1 + \frac{n}{\left[\sum_{i=1}^{n} \ln\left(\frac{x_i}{x_{\min}}\right)\right]}$$

Equation 4-5 Newman's power law exponent

Where x_i is the measured frequencies of x and x_{min} is the minimum *x* value for which the power law behaviour holds. The distributions in Figure 4-12 show that in all networks there is a disproportionate difference between authors with one and two papers that makes a slightly skewed rather than a straight curve. Therefore, the x_{min} is equal to 2, as the values after it present an almost linear behaviour.

GIScience Co-authorship Network	X_{min}	σ (using Equation 4.5)	σ ' (using Excel)
4-2	2	2.02	2.75 (2.81)
4-3	2	2.24	2.76 (2.66)
4-4	2	2.33	2.69 (2.55)
5-2	2	2.19	2.78 (2.59)
5-3	2	2.21	2.77 (2.64)
5-4	2	2.17	2.87 (2.65)

Table 4.15 Computed power law exponents (σ) for the Author Participation Distributions. The third column shows exponents calculated using x_{min} parameter and Equation 4-5. In the last column, σ ' is calculated using the least- square function implemented in Excel. In this column, the left hand value is the result of using all x- range and the value in parenthesis is the result of excluding values of x less than the x_{min} parameter.

The Table 4.15 shows the values of the exponents obtained (1) by using Equation 4-5, and (2) by extracting the slope of a fitted straight line in log-log plot using Excel (see Figure 4-12). The former values are similar to the *-2 exponent* were found by Alfred Lotka and expressed in his famous Lotka's Law of Scientific Productivity. The latter shows overestimated exponents (for the full x range and from $x > x_{min}$) as argued by Adamic. and Huberman (2000) and Newman (2005). One could say that the lower the exponent, the greater the inequality in the number of papers per authors.

Hence, one expects to find more inequalities in terms of paper per author numbers (more scholars with one paper and few with a large amount) in 4-2 than in 4-4, or in 5-2 than in 5-3 network. One can speculate that a very disperse topical coverage, such as the case of 4-2 or 5-2, may partially account for the difference in the number of authors with 1, 2 and 3 papers. It may be concluded that the more concentrated the network's coverage on GIScience or related topics, the more similar the authors' participation will be. However, it is important to note that according to Lotka's scientific productivity law, it is common in a co-authorship network to find a large number of authors with few papers. Nevertheless, the topology of a scientific collaboration network is not only shaped by the authors' network participation, but also by the size of co-authorship teams and group of authors' collaborators. These author patterns are examined in detail in the following two sections.

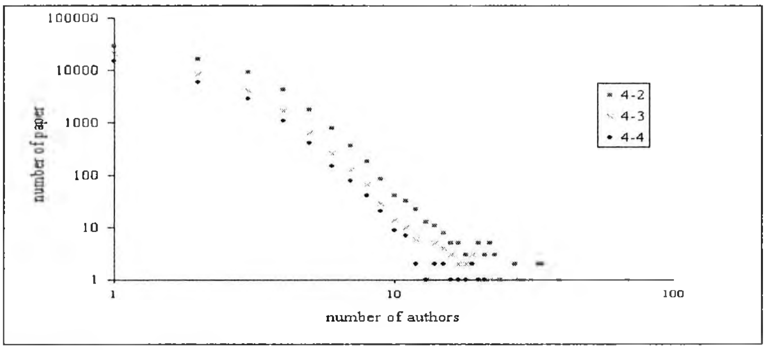
4.1.3 Co-Authorship Patterns in different GIScience Research Networks

Both, traditional bibliometric studies of scientific collaborations (Price, 1965; Melin and Persson, 1996; Logan *et al*, 1991; Ding *et al*, 1999; Otte and Rousseau, 2002), and more recent research on large-scale co-authorship networks (Newman, 2001a; Horn *et al*, 2004; Barabasi *et al*, 2004; Moody, 2004; Liu, *et al*, 2005; Wagner and Leydesdorff, 2005) have reported a steady increase in the number of collaborative papers in diverse fields, subfields and disciplines. As a result, later empirical studies have constructed relatively complete collaboration networks for a variety of disciplines and hence, are able to characterize their patterns of collaboration or co-authorship. Some scholars have compared the results, noticing that the patterns vary considerably from discipline to discipline (Grossman and Ion, 1995; Newman, 2004c; Moody, 2004), concluding that in some way, co-authorship patterns may identify real differences in the way research is carried out. For example, Moody (2004) states that in social sciences, quantitative work is more likely to be co-authored than non-quantitative work. Newman (2004c) confirmed that biologists tend to work in larger teams than mathematicians, among whom individual work still prevails. Nevertheless, an increasing tendency towards collaboration in this discipline has been detected (Grossman and Ion, 1995; Barabasi *et al*, 2004; Newman, 2004c).

Table 4.13 shows in the row "mean author per paper" the co-authorship patterns exhibited by different disciplines. According to the figures, mathematicians hold the lowest mean of 1.5 authors per paper. Grossman and Ion (1995) reported that during the forties, over 90% of papers in mathematics were works by sole authors. A higher co-authorship mean is exhibited in physics, a more theoretical than empirical discipline, where papers are written by teams of around 2.5 authors. However, Newman (2004c) emphasizes that the mean value varies substantially if the network is analyzed by subfields. For example, the co-authorship mean increases to more than 3 authors per paper in the astrophysics network, but decreases to less than 2 authors in high energy physics. Additionally, Moody (2004) found that social science is a discipline highly constrained by research speciality with a mean of 2.5 authors per paper, similar to physicists. However, from all networks compared, it is biomedical research that exhibits the highest mean with groups of around 4 authors writing papers.

Despite that 4-2 and 5-4 networks are two different representations of collaborations in GIScience, it seems that papers in both networks are typically works with around two authors (see Table 4.13). The average slightly decreases to 1.8 authors per paper for 4-3, 4-4 and 5-3 and keeps around the same for 5-2 network (see Table 4.12). Apparently, the results show that works by very large teams of authors are not common in any GIScience research network.

The results are better understood by examining the distribution of the number of authors per paper displayed in Figure 4-13.



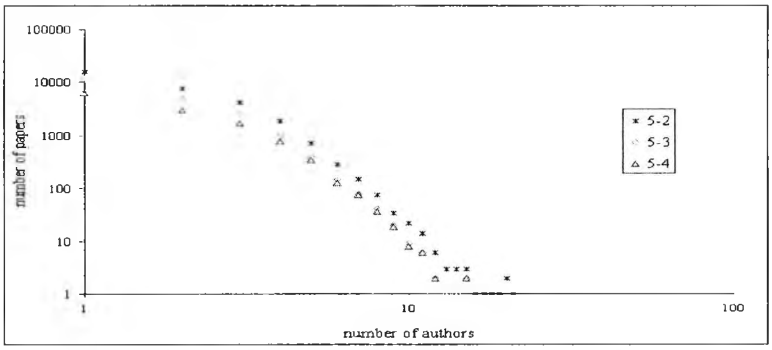


Figure 4-13 Distribution of the number of paper-authors in all GIScience networks (Log-Log Plot). The top panel shows the distribution for 4-X networks. Bottom Panel shows the distributions for 5-X Networks

Despite exhibiting a power law's signature of a large number of smaller events (papers with few authors) and very few of larger events (papers by larger teams), the distributions do not follow a power law distribution in full. Instead, similar to Newman (2004b)'s results, the distributions seem to follow a power law only in their middle sections (straight line sections in the log-log plot). Moreover, for papers with one or two authors, as is observed in Figure 4-13, the curves slightly skew but become very messy for teams of more than 20 authors (especially in 4-2 distribution). In this case, authors argue that the best fit is a power form with an exponential cut-off[2] (Amaral *et al*, 2000; Newman, 2001a). This distribution suggests an underlying degree that follows a power law but with imposed constraints, such as the tendency toward sole works in some disciplines like mathematics or humanities, that limit the maximum values of *x* (Newman, 2001a).

For all networks, it seems that the cut-off is around 10 authors per paper. Hence, one does not expect to find many papers written by teams larger than 10 authors in any of the networks. One can see a trend in 5-X series toward writing papers with very few

[2]A power law form with an exponential cut off is defined by $P(x) \sim x^{-\alpha} (e^{-x/x_c})$,where α – power law exponent and x_c - are constants. The calculation of these values is out of scope of this research.

others instead of with very large academic teams. Apparently, as topics get more relevant to GIScience, the academic teams that write the papers become much smaller. A fact that is reflected in a much lower co-authorship mean in 5-X networks than in 4-X. Hence, one can speculate that in this case, it is the nature of the research per se that limits the size of the co-authorship teams to a few members.

4.1.4 Collaboration Patterns

Collaboration patterns, defined as the mean number of co-authors, with whom scholars for a given discipline wrote papers during a specific time period, have been used to highlight differences in the way scholars collaborate across different disciplines. Of significance are the differences in collaboration patterns reported by Newman (2004c) on studying mathematicians', physicists' and biologists' networks. These three disciplines have a very different mode of doing science that may be captured in Newman's results. He observed that in mathematics, a largely theoretical discipline, the results show not only the less collaborative papers (see 4.1.3 Section), but also the less collaborative authors. Despite a reported increase in the mean number of authors per paper (Grossman and Ion, 1995), mathematics is still an individual science with a mean of around 4 collaborators per author. In contrast, in a more experimental discipline, such as biology, the predominance of larger teams of co-authors (see section 4.1.3) contributes to a much higher mean of around 18 collaborators per author. Besides, Newman (2004c) found that the mean of 10 collaborators per author in physics compared with the mathematics and biology results reflects the jointly theoretical and experimental nature of works in this discipline.

However, the collaboration mean for a discipline varies if it is analyzed by subfields or areas. The results from a study of three separate physics subfields (Newman, 2004c) suggest that work in theoretical fields such as high-energy physics has a lower collaboration mean of around 4 collaborators than work on partial or mainly experimental fields such as condensed matter or astrophysics with higher averages of around 6 and 15 collaborators, respectively. In general for all GIScience networks, the collaboration averages do not display any noticeable difference with each other.

As shown in Table 4.16, they range from 4.4 to 5.5 collaborators per author, very similar results to those obtained for computer science or HCI collaboration networks. Nevertheless, 4-2, the largest network in size, shows a slightly higher average than the others, probably as a result of much larger co-authorship teams (see section 4.1.3). As this network embraces all other networks' data sources, one may conclude that the difference in the collaboration averages is due to the 33 journals that only this network covers. Probably, co-authorship relations from works published in those journals (see table 3.4 chapter 3) hold a more collaborative author-profile than those published in the common set of journals.

Table 4.16 Collaboration Statistics in all GIScience Networks

Networks	4-2	4-3	4-4	5-2	5-3	5-4
Total Authors	65410	32348	22154	31678	19513	13950
Average Number of Collaborators per Author	5.51	4.83	4.70	4.76	4.46	4.79
% Collaborative authors	87%	80%	80%	85%	80%	83%
% authors 1 collaborator	21%	23%	24%	22%	23%	21%
% authors 2 collaborators	20%	20%	21%	21%	20%	20%
% authors more than 40 collaborators	0.30%	0.06%	0.07%	0.05%	0.03%	0.04%
% Non-Collaborative Authors	13%	20%	20%	15%	20%	17%

Concerning collaboration patterns, another interesting measure is the number of those who apparently prefer to publish individually. Within the specific case of GIScience, the non-collaborative authors represent between 13-17% in 4-2, 5-2 and 5-4 networks, increasing up to 20% of the total, in the remaining networks. It leaves more than 80% (depending on the network, see Table 4.16) of authors with at least one co-authorship link. However, more than 40% of the collaborative authors have a maximum of only two co-authors in contrast to less than 0.5% of authors that have more than 40 collaborators (see Table 4.16). The collaboration distribution in Figure 4-14 displays a large group of authors with very few collaborators, in contrast to a small group of very collaborative authors (in the right hand tail). Like the author participation and co-authorship distributions, the collaboration distribution follows an approximate straight line for the most part of its range in a log-log plot. However, the curve slightly skews for the low number of collaborators (left hand side) and becomes very erratic for authors with large number of collaborators (right hand side).

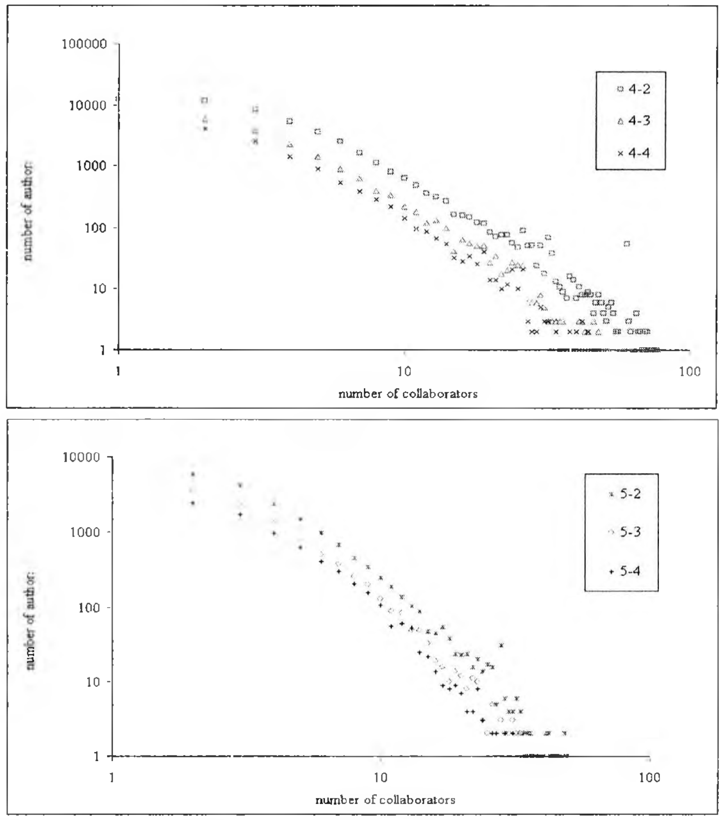


Figure 4-14 Histogram of the number of collaborators of authors in all networks (log-log plot). The collaboration distributions are calculated making no distinction between unique and new collaborators. (Unique collaborators' distribution or degree distribution is elaborated and analysed in section 4.3.1).

Therefore, authors studying collaboration distributions have suggested that either it can follow a power law with an exponential cut-off (Amaral *et al*, 2000; Newman, 2001a) or a power law that changes from one exponent (α) to another during its range (Barabasi, 2002; Newman, 2001b). As both mathematical analyses are out of the scope of this research, the distributions' exponents were calculated using the approach based on Equation 4-4 also deployed in 4.1.1. The following table shows for each network, the α results and the X_{min}, parameter used.

The first step to calculate the distribution exponent is to set up the X_{min} values. Unlike the authors participation distribution, X_{min} in this case takes a different values (see Table 4.17), depending on the network. It seems that the larger the network, the later in the x-axis it starts following a power law. Probably, it is due to the smaller difference between the authors with one and two collaborators that skews the curve at the beginning.

GIScience Co-authorship Network	X_{min}	σ (using Equation 4-5	σ ' (using excel)
4-2	6	2.02	2.73 (3.00)
4-3	4	2.02	2.81(3.02)
4-4	3	2.03	2.77(2.83)
5-2	6	1.93	2.89 (3.43)
5-3	5	1.99	2.83 (3.14)
5-4	3	1.92	2.85 (3.13)

Table 4.17 Power law exponents (σ) for the collaboration distributions. The σ is calculated using x_{min} parameter and Equation 4-5, and σ ' is calculated using the least- square function implemented in Excel. In this column, the left hand value is obtained by using all x- range, and the value in parenthesis is the result of discarding values of x less than the x_{min} parameter.

Regarding the power law distributions, Newman (2001a) found two different kinds of behaviour. On the one hand, for σ > 2, the average properties of the network are dominated by what he calls "little people" or individuals with few collaborators, whereas for σ < 2, networks are dominated by "hubs" or individuals with a large number of collaborators. These results show that while biomedical research network properties are dominated by those with fewer collaborators, in traditional science, network properties are determined by highly-connected scholars.

In GIScience networks, it seems that despite the presence of hub scholars in 4-X, they do not determine the average collaboration properties of these networks. In contrast, star-scholars seem to be more influential in 5-X networks. This hypothesis can be verified by calculating the degree of centralization for each network (refer to section 4.4). This section has separately analyzed network participation, and disclosed co-authorship trends and collaboration patterns. However, as each of these factors influence one another, they have to be examined altogether. Thus, one will be able to discern global topological features of the collaboration networks that can be utilized to understand the academic society of scholars working in GIScience research topics. In doing so, the following section examines topological features of the networks.

4.2 Topological Properties

In the previous section, basic statistical properties that describe collaboration and co-authorship patterns were analyzed. However, these statistical properties fail to inform about the structure of the co-authorship networks. Thus, this section focuses on the delineation and inspection of relevant topological properties that physicists and mathematicians (Abello *et al*, 1999; Newman, 2001a; Adamic and Huberman, 2002; Barabasi *et al*, 2004) have employed to analyze large network structures. In doing so, they have been able to characterize the topology of networks which are very dissimilar in nature such as the Internet, documents linked by citations, the World Wide Web, phone calls and protein folding. As a result, scholars have determined previously unseen characteristics and highlighted apparent or hidden differences and similarities among them. Within this section, the first sub-section compares and discusses key network concepts such as topological distance, geodesic path, largest component and diameter; computed for all GIScience collaboration networks. Based on these results the next sub-section outlines the differences and similarities between GIScience and other hard and soft disciplines.

4.2.1 Connectivity and Topological Distances in GIScience Co-authorship Networks

Section 2.2 in chapter 2 formally defines key network distances such as largest component, geodesic path and diameter, which are all based on the basic notion of topological distance. Thus, in a co-authorship network, the largest component corresponds to the largest interconnected scientific community, in which all authors can be reached through a finite number of co-authorship steps. In turn, the mean geodesic informs the average distance in terms of co-authorship links between scholars in the largest connected sub-community, or what is popularly named as the "degree of separation" (see section 2.2.1 chapter 2). Finally, the diameter corresponds to the maximum expected chain of intermediates between authors within the largest interconnected sub-community.

	GIScience Network					
	4-2	4-3	4-4	5-2	5-3	5-4
Total Number of Authors	65410	32348	22154	31678	19513	13950
% Authors connected in the largest sub-community	49%	36%	33%	36%	32%	31%
% Totally Isolated Authors (single disconnected nodes)	13%	20%	20%	15%	20%	16%
% Authors outside the largest component (but not isolated)	38%	45%	47%	49%	48%	52%

Table 4.18 Network distances and connectivity properties of six different GIScience co-authorship networks (calculated with Pajek Software[3]).

Table 4.18 shows the results of these three topological measures, revealing connectivity and distances patterns from published works extracted from different GIScience data sources.

[3] Free software for large network analysis and visualization, available at http://vlado.fmf.uni-lj.si/pub/networks/pajek/. Software's manual available at http://vlado.fmf.uni-lj.si/pub/networks/pajek/doc/pajekman.pdf.

4.2.1.1 Scientific Sub-Communities

In GIScience networks, the size of the largest components manages to fill only between 30% and 50% of the graphs. It reveals that via co-authorship only between 30%-50% of the authors are integrated, and so reachable from one another. However, it is surprising that 4-2, a network from a wide range of topics, rather than 5-4, a network embracing core GIScience topics, exhibits the highest sub-set of authors linked through co-authorship. For instance, the nature of the networks' data source per se may account for the unexpected situation. First of all, it is important to remember that the process of building each network involves the previous network data source and a new set of journals. Thus, there is an implicit order within the networks in each network series (4-X and 5-X), according to the nested data.

Table 4.19 shows the subsequent order and percentage of distinctive and shared journals comprising each network. The figures show that network 5-3 shares with 5-4 55% of its co-authorship links extracted from the common journals. In the same way, 5-2 has 37% unique collaboration links and shares 63% has 45% of unique co-authorship links.

Network	No Journals Data Source	distinctive journals	shared journals with the previous network
(5-4)	11	100%	
(5-3)	20	45%	55%
(5-2)	32	37%	63%
(4-4)	26	42%	58%
(4-3)	38	32%	68%
(4-2)	69	45%	55%

Table 4.19 GIScience Network and their shared data sources

Therefore, one can speculate that scientific collaborations extracted from around 30 journals (unique to 4-2 data source) may account for the 10% extra co-authorship links that connect almost 50% of authors in this network. Thus, network 4-2, despite its wide range of topics, manages to link the largest sub community of scholars. However, one cannot be sure that all collaborations extracted from these distinctive journals to 4-2 such as *Agriculture, Ecosystems & Environment, Catena, Cities, Computer Aided Design, Ecological Modelling, Regional Science or Urban Studies* are from works in GIScience or related topics. In the case they are, their authors are very likely to be linked through co-authorship to the largest group. Otherwise, they will be disconnected or part of smaller communities. The latter may be one of the reasons why there are approximately 30.000 disconnected authors in this network. In contrast, despite the fact that the 5-4 network covers core and very complementary research topics in GIScience, its co-authorship links only manage to connect 30% of the authors. One can conclude that, though working in GIScience issues, authors in this network are not very likely to collaborate between sub-networks, or may work on topics, that although similar, are not complementary.

Consequently, it is important to examine authors outside the largest component because they may be disconnected, but part of a significant sub-network that may be related to GIScience or corresponding to an allied discipline. Figure 4-15 reveals there are two types of authors outside the largest community. One comprises completely isolated authors (white portion) that participate in the network only through individual works (they do not collaborate). The figure shows that 4-2 exhibits the smaller proportion of the disconnected authors, despite of doubling in size to the smaller network. The other group comprises authors who, though disconnected from the largest component, are working as part of much smaller groups.

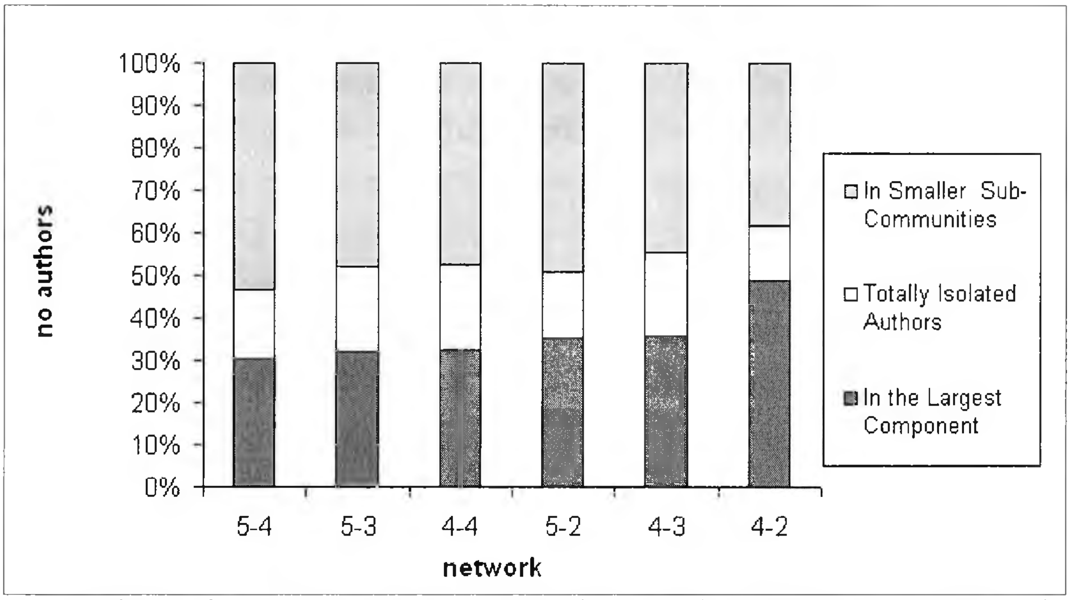


Figure 4-15 Sizes of the largest sub-communities (dark grey), smaller sub-communities (light grey) and totally isolated authors (white) for all GIScience networks studied.

Table 4.20 presents the sizes of these secondary communities in all networks. The figures shows that disconnected authors are not forming any significant (in size) sub-community. Therefore, one can conclude that the connectivity patterns for all networks comprise a giant component, filling a large proportion of the graph, accompanied with a large number of much smaller groups.

One can speculate that the largest disconnected groups may represent two different types of scholars' research backgrounds. On the one hand, scholars who work on totally un-related topics to GIScience, and therefore very unlikely to be part of the largest community. On the other hand, scholars that work on very complementary topics to GIScience, but due to the poor coverage of the network' sources appear disconnected.

Ranking	Co-authorship Sub-Communities (in number of authors)					
	5-4	5-3	5-2	4-4	4-3	4-2
Second	150	70	53	73	44	41
Third	62	40	39	50	36	40
Fourth	57	35	37	37	30	37
Fifth	41	30	34	35	29	33
Communities with more than 10 authors	78	84	186	82	127	181

Table 4.20 Sizes of the significant communities outside the largest component in all networks. Last row shows the number of sub-communities with more than 10 authors.

Specifically, considering the sizes of the second largest communities (see Table 4.20), it is evident that the 150 author group in 5-4 or 73 author group in 4-4 are linked to the largest component as more journals (probability very related to the authors topics) are added to the subsequent networks' data sources, 5-5 or 4-3 respectively. The same may occur in the other networks, as the sizes of the largest sub-communities (3rd, 4th and 5th) decrease as more journals are added to build the larger subsequent networks. Moreover, at the same time the larger sub-communities join the core of the networks, the number of smaller groups with more than 10 authors increases. Probably, it is a consequence of the addition of new journals that may bring works that are not related to the topics covered by the previous networks, hence, portraying authors within disconnected groups.

In summary, it is clear that outside the main components all networks are highly fragmented into groups. They may be working on specific themes that, either participate with few papers or only collaborate among themselves, resulting in disconnected islands with recurrent links. For example, the second largest sub-community in 5-4 network refers to 150 scholars working on remote sensing issues, mainly in India. In the same way, in 5-3, the third largest sub-community comprises 40 authors linked by publication on geomorphology, environmental and remote sensing issues. Moreover, the sub-community with 37 authors is fourth largest in the 4-4 network. In 5-2, the largest second component comprises collaborations from 53 authors on geomorphology, GIS applications, atmospheric science and remote sensing. Similarly, in 4-2, the second largest is a group linked by co-authorship connections on agriculture, water and remote sensing. Moving to less related topics to GIScience, in 4-3, the fourth largest community is linked by co-authorships on simulations and computer graphics.

At this level, one cannot say that the authors in these sub-communities are part of established research groups. What one can say is that they are linked by co-authorship due to similar research interests. However, for all networks, low network connectivity can be a consequence of either omissions of other important data sources very relevant to GIScience, or the presence of many journals not very relevant to GIScience. This, in turn, may be the cause of the low levels of network participation reported on section 4.1.2 of this chapter.

4.2.1.2 Degrees of Separation in GIScience

Typical average distances within a co-authorship network are the result of computing all authors' direct connections. If the co-authorship network comprises very collaborative authors then it is likely that the network exhibits small distances (mean geodesic) between a given pair of authors. Otherwise, one needs to get through more intermediates to reach one node from another.

	4-2	**4-2**	**4-4**	**5-2**	**5-3**	**5-4**
Mean geodesic (average distance or degree of separation)	8.75	10.37	10.44	10.25	10.58	9.52
Diameter	28	33	34	30	32	28

Table 4.21 Significant topological network distances within all six GIScience co-authorship Networks.

In the case of GIScience networks, though exhibiting low averages of collaboration, participation and co-authorship, distances are rather small. Table 4.21 shows that in all networks, the optimal distances or mean geodesic (mean geodesic were calculated only using authors in the largest component.) are around 9 – 11 co-authorship links. It is important to note that despite the difference in number of authors and connections among all networks, there is not a significant variation in the geodesic distances (see Figure 4-16).

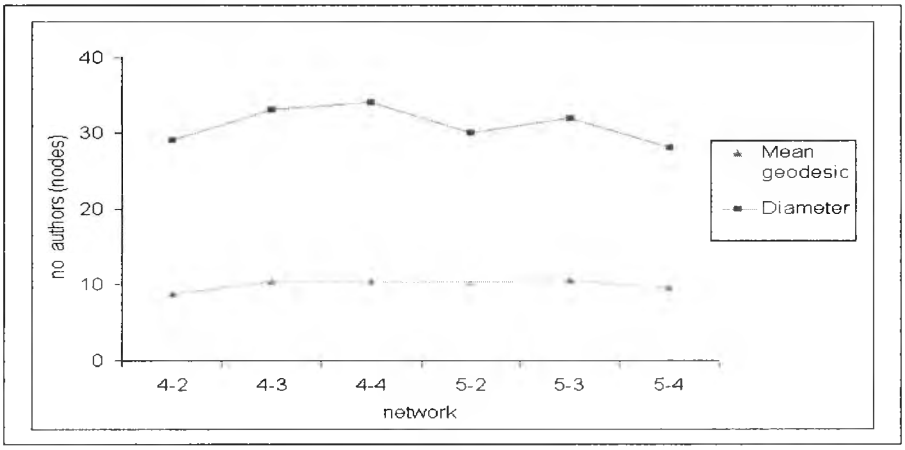


Figure 4-16 Comparing mean geodesic and diameter among all networks

Despite the similarities in the mean geodesic measures, it is necessary to take into account the difference in size of all networks. A chain of around 9 intermediaries becomes more meaningful in a network like 4-2 that is double the size of the others. It implies that the 4-2 network is more densely connected, and it would be quicker to reach any author. Hence, in the hypothetical case of transmitting information using 4-2 network structure, it not only implies that the information packet would reach a high number of scholars, but also, it may reach them quicker as less intermediaries are needed.

The most distant authors are separated by no more than 34 links in all networks. The diameters vary from 28 steps in 5-4 to 34 steps in 4-4 network. In 4-2, the furthest pair of authors *Brodlie, KW*, and *Streeton, CL* are a 28 intermediary chain apart. If this diameter is compared to the 30000 authors within the largest sub-community, 28 becomes a more meaningful distance. Figure 4-17 shows the complete chain of 28 intermediaries between this pair of authors. Recalling that common interests exist between any adjacent links, one can say that the diameter also provides a sequence of working research topics for the authors along it. As the diameter shows the more distant connected authors are very likely to work on dissimilar research topics.

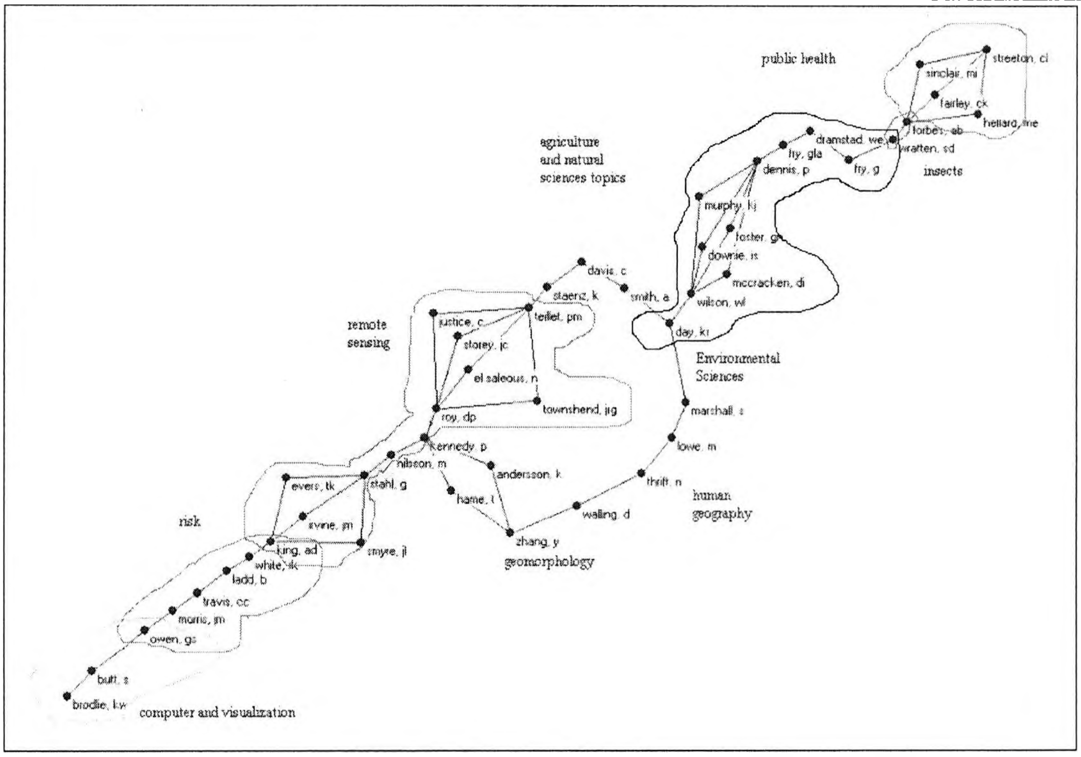


Figure 4-17 The 4-2 Network Diameter – 28 degrees from Brodlie, KW to Streeton, CL. Authors with similar research topics were manually circled. Thus, circles broadly represent all research areas along the diameter.

Broadly, Figure 4-17 shows authors, manually circled, that apparently work on similar research topics. The sequence starts with *Brodlie, KW* (panel's left-bottom corner) working on computing visualization, then, it moves on to a large group working on remote sensing. As there is a bifurcation, left branch authors also work on remote sensing, while, authors on the right branch works geomorphology and human geography. They converge on environmental and natural topics, finally to end in *Streeton,* CL's work on public health medicine. The diversity of topics within the diameter is a consequence of a wide range of journals' topics that 4-2 embraces. However, it is also clear, for example, that certain topics such as geomorphology, public health or risk are less relevant to GIScience, compared to remote sensing. Nevertheless, GIScience acts as a direct or indirect commonality that links, via co-authorship, authors working on a variety of topics.

In contrast, the high relevance to GIScience is noticeable for the majority of the topics along the 5-4 diameter chain, shown in Figure 4-18. The figure shows that the topics are mainly concerned with diverse applications of remote sensing.

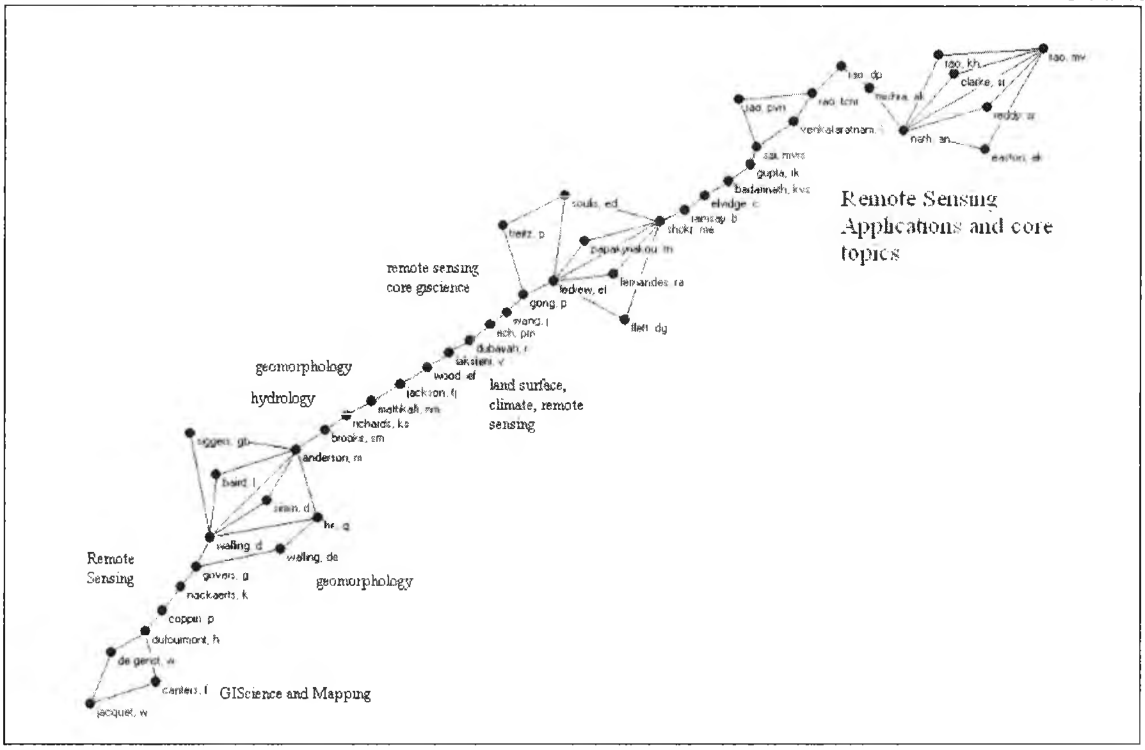


Figure 4-18 The 5-4 Network Diameter and its adjacent topics- 28 Degrees from Jacquest, W to Rao, MV. Authors with similar research topics were manually circled. Thus, circles broadly represent all research areas along the diameter

They start with works on GIScience and mapping. Then, they shift to authors working on diverse applications of remote sensing that involve geomorphology, hydrology and climate. Unlike 4-2, all topics adjacent to 5-4 diameter are from remote sensing, a strongly allied GIScience discipline.

4.2.2 GIScience compared to other domains

Empirical studies on the structural properties of co-authorship networks extracted from large bibliographic databases have helped to characterize entire disciplines, fields and areas (Newman, 2001a; Horn *et al*, 2004; Newman, 2004c; Moody, 2004). Therefore, the purpose of this section is to compare distance and connectivity patterns in GIScience networks to those obtained from other disciplines or research domains.

- Firstly, to newer research domains or subfields such as Computer Science or HCI.
- Secondly, to more traditional or well-established disciplines such as Mathematics or Physics.
- Thirdly, to special cases such as Biomedical research that, according to Newman (2004c), account for the largest part of civilian scientific research by far.
- Finally, to Social Science patterns, as GIScience is a branch of information science, which in turn can be considered in the Social Science domain.

In doing these comparisons, one can establish similarities or dissimilarities between GIScience and other disciplines which are very different in nature.

In this case, research in GIScience is represented by 4-2 and 5-4 collaboration networks. On the one hand, 4-2 is the result of collaborations in a wide range of GIScience topics that cover a very broad research domain. On the other hand, 5-4 depicts a collaboration network that embraces closer, but at the same time, highly dissimilar research topics to GIScience.

Property	**5-4 GI Science**	**4-2 GI Science**	**Computer Science-preprints**	**HCI**	**Bio-medical Research**	**Physics (online preprint database)**	**Mathematics**	**Social Science**
Largest Component Size	31%	49%	57%	51%	92%	85%	82%	68%
Mean Geodesic	10.0	9.0	9.7	6.8	4.6	5.9	7.6	9.8
Diameter	28	28	31	27	24	20	27	na
Time Window	1992-02	1992-02	1995-99	1998-02	1995-99	1995-99	1995-99	1969-99
Work's Reference			(Newman, 2001a).	(Horn *et al*, 2004)	(Newman, 2004c).	(Newman, 2004c	(Newman, 2004c	(Moody, 2004)

Table 4.22 Summary of the main topological network properties of eight co-authorship networks. Two from GIScience and five from other fields, sub-fields and disciplines.

Results displayed in Table 4.22 show that biomedical research exhibits the highest level of connectivity among all networks. Newman (2004c) reports that around 92% of authors are reachable from one another. It is not a surprising result, as on average, a biomedical author is linked to around 18 others and research teams comprise around 7 authors. Higher network connectivity is also exhibited by mathematicians, with around 80% of the authors inter-connected despite smaller research teams (less than two authors on average). Physics also exhibits a higher degree of connectivity with its largest component filling the 85% of the network; probably as the result of larger research teams of around 10 collaborators.

In contrast, the level of connectivity of network 4-2, slightly less than 50%, is very similar to computer science or HCI. A lower level is exhibited by the 5-4 network, where only around 30% of the authors are connected to the largest common network by co-authorship links. It is also the lowest connectivity level among all GIScience networks. One possible explanation is related to the number and nature of research topics covered. Though very related to GIScience, the topics in the 5-4 data source cover independent GIScience research lines with authors not doing cross-topic collaborations (see section 4.1). Besides, there is also a limit to the coverage regarding the number of journals from each topic. Thus, the combinations of these factors may account for the most disconnected network with the lowest authors' participation.

The largest component sizes depend not only on the time window mapped, but also on the nature of the scientific documents used to build the network e.g. pre-prints, refereed works, books, etc. They also depend on the exhaustiveness of the data source representing the field or discipline literature. For example, in the largest social science component only 68% of the authors are connected, despite covering a 30 year time period. In contrast, for a smaller time window, the physics network shows much higher connectivity levels. However, this network is built up from pre-prints instead of refereed works. This may have helped to increase authors' participation as all works are accepted without further examination.

As expected, global distances in the biomedical network are much shorter than in networks with lesser connected cores such as GIScience, HCI, social science and computer science. While an average distance between a pair of biomedical scholars is less than five intermediaries, it increases to around ten for both GIScience and computer science networks. Furthermore, biomedical distances are even shorter than in similar interconnected networks such as mathematics and physics where a pair of authors is separated by five to eight steps.

In summary, all GIScience collaboration networks exhibit similar distance patterns where short average distances predominate. Besides, all networks comprise a large connected community with many much smaller groups of not very significant sizes. However, the networks' connectivity levels vary from 30% to around 50% of the total number of authors. As all networks were built in a similar way, they may be the extra connections from one data input to another that may account for the differences in the networks' connectivity levels. In turn, the lowest connectivity displayed by network 5-2 maybe the result of mapping only research in core GIScience that according to the results, displays a low tendency towards collaboration crossing speciality borders. In contrast, 4-2, made up from wider topic coverage, shows a larger connected community.

Overall, the results indicate that it is more likely that any pair of authors is connected by co-authorship links within a wide coverage network than within a network that only covers core GIScience topics. In the case of 5-4, one may argue that lower coverage could be translated into the presence of smaller disconnected sub-communities of significant sizes. However, as the size of the second largest sub community indicates, none of the network topologies indicates the presence of relevant isolated groups surrounding particular topics. However, as the 5-4 network's collaborations are embedded in 4-2, it is certain that there are vital collaborations that bridge the 5-4 network gaps; resulting in a more interconnected community. Thus, it is crucial to identify the authors that do not only collaborate with fellow colleagues, but with scholars from other fields. The following section considers those who are playing critical and central roles within the networks.

4.3 The Core of GIScience Collaboration Networks

The results of this research and other similar works, presented in sections 4.1 and 4.2, highlighted inequalities regarding authors' participation, co-authorship and collaboration between not only disciplines but also between scholars within each discipline. In the specific case of GIScience, these inequalities are easily detected by examining the frequency distribution plots of these measures (see Figure 4-12, Figure 4-11 and Figure 4-14). Thus, it is evident that there are few papers by large teams, and few highly participative or collaborative authors. Therefore, one may expect that within the network topology, these highly participative or collaborative authors hold more influential or central positions than others.

One of the underlying premises of SNA is that a global network structure is the result of actors' local interactions (Otte and Rousseau, 2002). Therefore, SNA offers mechanisms not only to study the global properties of relational network data, but also, to evaluate actors' locations and their influences. Within a communication network structure *central actors* are regarded as having better access to and control over information (Peterson, 1998; Albert and Barabasi, 2002; Newman, 2003). Moreover, Moody (2004) argues that within a collaboration network, a central position regarding a large number of collaborators helps central authors to more rapidly diffuse their ideas than authors with an average number of collaborators. Therefore, in SNA context central authors are often defined by

- The ease of communicating with many others
- The proximity to many others
- And the frequency of their role as intermediary in the other nodes' interactions (Freeman, 1979; Otte and Rousseau, 2002; De Nooy *et al*, 2005; Elmacioglu and Lee, 2005).

Consequently, this section concentrates on central authors and the relevance of their connections within the shaping of GIScience network structures. In doing so, it deploys a widely used SNA ego-centred approach (Newman, 2001a; Otte and Rousseau, 2002; Horn *et al*, 2004; Moody, 2004; Elmacioglu and Lee, 2005) based on degree, closeness and betweenness centrality measures defined by Freeman(1979).

Freeman (1979)'s measures are the most popular SNA centralization measures implemented in many SNA software packages (Ucinet, Pajek or Netdraw). Though independent of the network' size, they rely only on geodesic distances (Cornwell, 2005). Hence, the measures intrinsically assume that communication only takes place along the shortest paths in the network; neglecting the possibility that it may take longer routes (Stephenson, and Zelen,1989; Borgatti, 2005; Cornwell, 2005; Newman, 2005a). Moreover, they are only defined for connected networks, thus networks outside the largest component are not taken into account (Cornwell, 2005; Newman, 2005a). Hence, Freeman's centrality measures are not suitable for the networks where information does not flow along the shortest paths, or where it is important to take into account nodes outside the largest component.

Nevertheless, Freeman's measures were used, first, because they allow comparison of the six GIScience networks despite having different sizes. Second, even though they do not take into account disconnected nodes, they allow analysis of the core of the networks where authors are connected via GIScience co-authorship links. Finally, this research does not intend to evaluate the flow of information in the co-authorship network, but rather to identify central authors and to speculate on the benefit of their privileged positions at the time to spread ideas or findings. Thus in the context of this research, Freeman's measures are very useful.

4.3.1 Degree Centrality

In graph theory, the *degree centrality* of a vertex is its degree or number of edges incident within it (see Chapter 2, section 2.2 for a mathematical definition). This is the simplest concept of centrality based on the premise that locally well-connected actors can more easily access the resources available within the network than actors with fewer connections (Freeman, 1979). Hagen *et al*, (1997) interprets node degree as a measure of a given node network activity; the more links a node has the more it can be regarded as active.

When studying the structure of a co-authorship network, the node degree reveals the number of co-authorship connections of a given author. However, as Newman (2003) explains, within a co-authorship network the node degree is not necessarily equal to the number of neighbours or collaborators to an author. Since a pair of authors can collaborate multiple times, there may be more than one link between them. Nevertheless, De Nooy *et al* (2005) propose representing a co-authorship network as a simple directed graph in which redundant collaboration between authors are added up and interpreted as the strength of a tie. In doing so, the number of collaborators (analyzed in section 4.1.4) becomes equal to node degree. The degree distribution, examined in Figure 4-14, evaluates the level of activity within the networks. Likewise, it makes evident a very small group of highly collaborative authors in contrast to, a large number of authors that exhibit few.

Examining the degree ranking in Table 4.23, three important groups were identified and highlighted with different colour shades. The first group, shaded in light green, comprises C. Justice the only author ranked in all networks. As his number of collaborators increases from 53 in 5-4 to 83 in the 4-2 network, one can say that his works are not only on relevant GIScience topics covered by the 5-4 network, but also from works on more general topics covered by 4-2.

The second special group, coloured in grey, comprises those top authors from whom the number of collaborators appears approximately constant through all networks (except 4-2). Hence, their connections are the result of works published in the 11 journals covered by 5-4, which are the core of all other networks' data sources. Moreover, as 5-4 network data source is identified as from allied or core topics to GIScience, one can regard these top-authors as scholars working in GIScience or very strongly related research topics. However, their degrees do not classify them in the top rank in network 4-2. Moreover, apart from C. Justice there are no authors form other rankings in network 4-2.

	4-2	
	authors	**node degree**
1	williams, r	99
2	running, sw	87
3	jenkins, a	86
4	cooper, c	84
5	bessemoulin, p	83
6	justice, c	83
7	smith, gd	82
8	nilsson, b	81
9	ziegler, r	81
10	privette, jl	79
11	nelder, ja	78
12	lindley, dv	76
13	sellers, p	76
14	barnard, ga	75
15	gash, jhc	75
16	jones, pd	75
17	kabat, p	75

4-3	
authors	**node degree**
nelder, ja	78
lindley, dv	76
barnard, ga	75
bartholomew, dj	54
justice, c	53
hand, dj	50
cracknell, ap	47
preece, da	47
cox, dr	45
curran, p	45
govers, g	45
zech, w	45
box, g	44
chatfield, c	43
gong, p	42
goodchild, m	42
jensen, jr	42

4-4	
authors	**node degree**
lindley, dv	76
barnard, ga	75
nelder, ja	62
bartholomew, dj	54
justice, c	53
hand, dj	50
cracknell, ap	47
preece, da	47
cox, dr	45
curran, p	45
chatfield, c	43
gong, p	42
goodchild, m	42
jensen, jr	42
evans, k	41
longford, nt	40
navalgund, rr	39
townshend, jrg	39

5-2	
authors	**node degree**
jenkins, a	74
justice, c	55
zech, w	55
olsson, m	48
cracknell, ap	47
richards, ks	47
stein, a	46
curran, p	45
tranter, m	44
goodchild, m	43
gong, p	42
jensen, jr	42

5-3	
author	**node degree**
justice, c	53
cracknell, ap	47
curran, p	45
gong, p	42
jensen, jr	42
goodchild, m	41
navalgund, rr	39
townshend, jrg	39
govers, g	38
jackson, tj	38
karnieli, a	38

5-4	
authors	**node degree**
justice, c	53
cracknell, ap	47
curran, p	44
gong, p	42
jensen, jr	42
goodchild, m	39
navalgund, rr	39
townshend, jrg	39
govers, g	38
jackson, tj	38
karnieli, a	38

Table 4.1 Node degree top-ranking for all networks. Authors are shaded according to their presence (high degree rank) throughout all networks. In light green is the only author ranked in all networks. In contrast, pink authors are highly connected authors only in network 4-2. Grey authors represent those who are ranked in all networks but network 4-2

Authors ranking in 4-2, shaded in pink, comprise the special third group. They are authors whose their collaborations only classify them in this ranking. The reason why it happens is that node degree ranking in 4-2 is heavily influenced by the presence of a paper co-authored by a rather large group of 68 authors. The paper was published in the *Journal of Epidemiology and Community Health* in 1995. As a result, co-authors of this paper exhibit a minimum degree of 67 collaborators, plus all links from all other additional papers they wrote. Before 1995, the highest node degree in this network was 32 collaborators exhibited by M. Goodchild. After 1995, R. William becomes the author with the highest node degree, as the result of being one of the co-authors of the abovementioned paper. From this cohort of 68, there are 14 other authors placed in the degree top-30 rank. However, neither of them is ranked in any of the other networks.

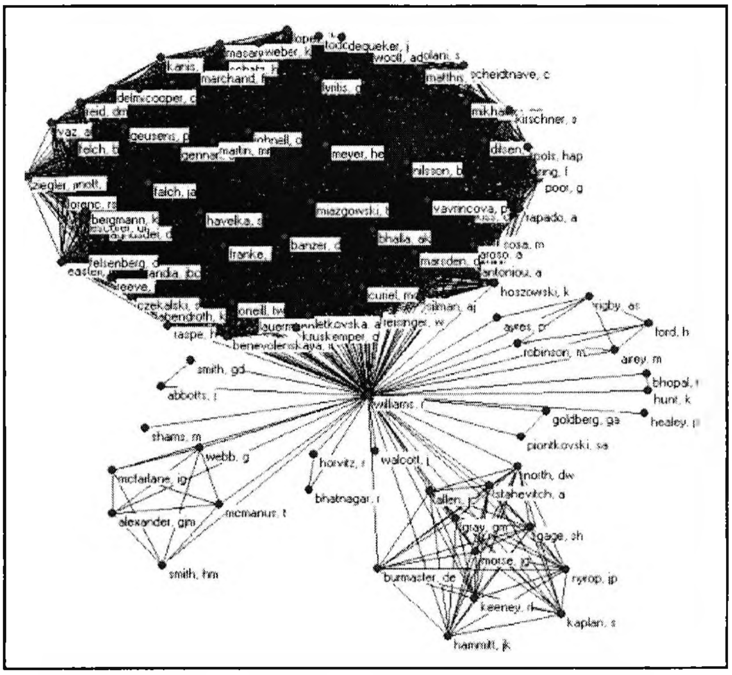


Figure 4-19 Egocentric network of R. William, the highest node degree author in all networks.

Figure 4-19 shows R. William's collaboration network; in just one step he can reach 99 other authors within the 4-2 network structure. However, his node degree decreases to only 5 collaborators in network 4-3, as the following networks become more restricted in terms of journal topics. Hence, they do not cover the author publication sources. In the figure, the highly connected sub-network in the top left corner of the panel corresponds to the 68 author paper and the bottom part to the other papers by this author. From this group, there are 53 authors with only 67 collaborators, suggesting that these authors contribute only with a paper in a medical journal not so relevant to GIScience. Moreover, none of their works are included in any subsequent network, as less related journals to GIScience are excluded.

Similarly to R. Williams is the case of A. Jenkins who exhibits a high node degree in 4-2 and the highest in 5-2, but, does not appear in any of the subsequent 4-3, 4-4, 5-3 or 5-4 networks' rankings. However, there is a crucial difference between these two authors, R. Williams published in journals that are not very relevant to GIScience. In the case of A. Jenkins, though he also published two highly collaborative works (with 16 and 21 co-authors), not only these two, but the majority of his works are published in journals such as *Geography*[4] or *Journal of Geography In Higher Education*[5] that cover relevant topics to GIScience. Hence unlike J. Williams[6], one expects that Jenkins does research in GIScience or relevant topics. Therefore, though both authors have a very high number of collaborators, due to the nature of their collaboration per se, they will occupy very different positions within the GIScience global network topologies. One can conclude that as each network data source becomes more selective about a journal's relevance to GIScience, more well known authors in the discipline start appearing in the degree top ranking list (Table 4.23).

Finally, the degree rank identifies very active authors regarding their immediate connections within each network structure. However, one cannot say to what extent being very active places authors in an advantageous structural position within the network topology. In order to answer this question, authors' positions regarding their betweenness and closeness centrality measures are analyzed in the next sections.

4.3.2 Closeness

Knowing only how many others a node is connected to does not suffice to characterize the network's structure or to evaluate the relevance of the nodes' positions within the overall network. Freeman (1979) proposed a centrality measure based upon the degree to which a node is close to all other nodes in the network. If a path between two nodes exists, the shorter the *distance* the closer they are. Hence, the easier the exchange of information would be with fewer message transmissions, shorter times and lower costs (Freeman ,1979). Set into the context of academic networks, shorter distances between scholars may benefit science as, according to Newman (2001b), scientific information such as discoveries, experimental results or new theories do not have to travel through the whole network of scientific acquaintances to reach those who can benefit by them.

[4] Paper: The Impact Of The Research Assessment Exercises On Teaching In Selected Geography - Departments In England And Wales.

[5] Paper: Moving with the Times: An Oral History of a Geography Department.

[6] Papers: Association of medical, physiological, behavioural and socio-economic factors with elevated mortality in men of Irish heritage in West Scotland and Principles for conduct of pest risk analyses: Report of an expert workshop in Risk Analysis Journal.

Following Freeman's (1979) definition, *closeness centrality* of a node calculates the total distance (number of paths) of this node from all other nodes. In a co-authorship network, authors with low closeness scores are regarded as those who may access quicker new information, or as producers, who may disseminate quicker information to others. Hence, a smaller closeness value means a most central author regarding information availability, or in general, network resources accessibility (Newman, 2001b). However, Freeman's closeness centrality is not well-defined in a weakly connected network (Liu, *et al*, 2005) as it requires the existence of paths connecting all nodes. Consequently, the measure was only applied to nodes within the largest component of each network.

	4-2	
	Author	**Closeness ($x\ 10^1$)**
1	sellers, p	1.76
2	running, sw	1.74
3	**justice, c**	1.73
4	melack, jm	1.70
5	kabat, p	1.70
6	privette, jl	1.69
7	goward, sn	1.68
8	meybeck, m	1.68
9	zhang, y	1.68
10	myneni, rb	1.68
23	**chen, j**	1.65
38	**openshaw, s**	1.64
115	goodchild, m	1.59

4-3	
Author	**Closeness ($x10$)**
openshaw, s	1.51
goodchild, m	1.51
foody, g	1.48
martin, d	1.46
krug, t	1.46
dorling, d	1.46
martin, r	1.46
brunsdon, c	1.44
flowerdew, r	1.44
chen, j	1.44
justice, c	1.40

4-4	
Author	**Closeness ($x10^1$)**
openshaw, s	1.53
goodchild, m	1.53
chen, j	1.49
estes, j	1.48
krug, t	1.47
martin, r	1.47
li, zl	1.47
justice, c	1.47
flowerdew, r	1.47
wise, s	1.46

	5-2	
	Author	**Closeness ($x10^1$)**
1	wang, j	1.53
2	curran, p	1.49
3	**goodchild, m**	1.48
4	bastiaanssen, wgm	1.48
5	foody, g	1.48
6	atkinson, p	1.47
7	**openshaw, s**	1.47
8	li, zl	1.46
9	**chen, j**	1.46
10	stein, a	1.45
19	**justice, c**	1.43

5-3	
Author	**Closeness ($x10^1$)**
justice, c	1.49
goodchild, m	1.48
openshaw, s	1.48
townshend, jrg	1.47
belward, a	1.46
chen, j	1.45
li, zl	1.45
estes, j	1.45
krug, t	1.43
gong, p	1.43

5-4	
Author	**Closeness ($x10^1$)**
justice, c	1.67
belward, a	1.66
townshend, jrg	1.63
estes, j	1.62
openshaw, s	1.61
goodchild, m	1.61
teillet, p	1.59
malingreau, jp	1.59
chen, j	1.59
li, zl	1.58

Table 4.24 Top-ranking closeness centrality in the largest components of all networks. Authors in bold hold up their central position through all networks.

Table 4.24 shows authors ranked by their centrality scores (calculated using Pajek software) in each studied network. Some conclusions can be drawn from top-ranking authors regarding access to information. There are four authors, in bold , C. Justice, M. Goodchild, S. Openshaw and J. Chen who hold up consistently their central positions through all network rankings. Manually inspecting their academic backgrounds, one realises that they work on very close topics to GIScience. For example, C. Justice works on remote sensing issues, M. Goodchild on various GIScience core topics, S. Openshaw, retired in 2000, mainly worked on geo-computation and J. M. Chen (Jing Ming Chen) works on GIS and remote sensing applications. Additionally, they are also active scholars as they have worked with far more colleagues than average collaborators in each network. Despite that, only C. Justice is consistently ranked in the all networks in Table 4.23.

Therefore, within the academic society represented by the co-authorship network, one can speculate that these are scholars who may exert strong influence due to their structural positions. On the one hand, to be in contact with many other colleagues with whom they collaborate, affords them many alternative communication paths. Hence, they become less dependent on others than scholars with less collaborators. As there are less intermediaries between them and all others in the network (high closeness scores), their ideas, opinions or findings can spread faster, but, crucial information or new developments also reach them quicker. Furthermore, they are not only close to many in the network, but they exhibit short geodesic distances among themselves despite working on different issues in GIScience. For example, Justice and Goodchild are two degrees apart linked by a mutual collaborator, the late Pr. John E. Estes, a pioneer in the fundamental and applied aspects of remote sensing and geographic information. Justice and Openshaw are 3 degrees apart, Openshaw and Goodchild only one degree, and there are three degrees between Chen and Goodchild.

Additionally, Table 4.24 shows a second group of scholars such as JHC. Gash, A. Jenkins, J. Estes and T. Krug (in italics) that despite not holding a high closeness scores in 4-2, do exhibit highly closeness scores in the subsequent networks. It shows that their connections placed them close to scholars covering topics more related to GIScience. However, none of the 68 authors in the paper mentioned in section 4.2.1 despite their rather high node degree exhibits a high closeness score. Not even R. Williams whom is placed in 1168th of the closeness rank.

Finally, in 4-2, authors such as SW. Running, C. Justice, JI. Privette, P. Sellers and P. Kabat are not only very participative (in the top-20 node degree ranking), but also they hold a very structurally relevant place in the network topology to be close to many.

However among them, only C. Justice appears in the top closeness ranking in network 5-4. Hence, one can speculate that those authors may be very active and locally well connected to scholars who like them, publish only in journals that are part of the 4-2 network.

4.3.3 Betweenness

Freeman (1979) also proposed a measure of centrality based upon the frequency with which a point falls between pairs of other points on the shortest or geodesic paths connecting them. Hence, he defines betweenness centrality as the proportion of all geodesics between pairs of other vertices that include the given vertices. In other words, this figure measures the extent to which a node facilitates the information flow in a communication network. Freeman (1979) advocates that the more a person is strategically located on the shorter communication paths linking pairs of others, the more central he/she becomes. This structural advantage has been understood in different ways. Newman (2004b) explains that nodes with high betweenness scores play the role of connecting different groups, thus, one can regard them as acting as a broker within the network (Otte and Rousseau, 2002). Moreover, Elmacioglu and Lee (2005) state that their role becomes more crucial, as they are part of the paths that lead to the fastest interactions of two non-directly connected authors. Thus, according to Liu *et al* (2005), betweenness can be used as a measure of the influence a node has over the spread of information through the network.

Table 4.25 shows the highest authors' betweenness scores for all networks. Firstly, it is important to note that not all authors lie in between others shortest paths. In fact, in all GIScience networks approximately 70% of authors in the largest component score a betweenness value of zero. It leaves only 30% of authors with betweenness scores greater than zero, in a position of hypothetically controlling the information flow within the networks. Analyzing Table 4.25, there is a marked difference among the highest betweenness scores in all networks. As the networks become smaller in size and their scope limited to more relevant journals in GIScience (5-4 or 4-4), the highest betweenness scores increase. For example, top scores' authors in 4-2 are in-between 3-4% of the paths, the scores increase to 5-9% in 4-3, and to 7-12% of the shortest paths in 4-4. In the same manner, in 5-2, top authors lie in 5-9-% of the shortest paths, while the percentage increases to 9-15% in 5-3, and finally, top authors lie in 8-16% of the shortest paths in the 5-4 network. Thus the relationship between the size of the network, translated into the number and topics of journals included, and the betweenness centrality scores of the top authors is evident. For instance, those who score higher betweenness in network 5-4 are in-between more scholars than those with higher betweenness in network 4-2.

	4-2			4-3			4-4	
	Author	**Betweenness (x 10^2)**		**Author**	**Betweenness (x 10^2)**		**Author**	**Betweenness (x 10^2)**
1	kim, j	4.36		**goodchild, m**	9.69		**goodchild, m**	12.04
2	running, sw	4.15		martin, d	9.60		**openshaw, s**	10.83
3	**openshaw, s**	3.96		foody, g	8.83		foody, g	9.86
4	jenkins, a	3.73		**openshaw, s**	7.90		gong, p	9.23
5	**justice, c**	3.40		curran, p	6.34		fuller, r	8.88
6	**chen, j**	3.11		dorling, d	6.01		**justice, c**	8.25
7	martin, d	2.91		gong, p	6.00		elvidge, c	7.94
8	smith, gd	2.90		**chen, j**	5.50		white, k	7.90
9	li, zl	2.68		elvidge, c	5.31		wang, j	7.75
10	**goodchild, m**	2.68		li, zl	5.23		gupta, rk	7.46
			12	**justice, c**	5.11	14	**chen, j**	7.14

	5-2			5-3			5-4	
	Author	**Betweenness (x 10^2)**		**Author**	**Betweenness (x 10^2)**		**Author**	**Betweenness (x 10^2)**
1	**goodchild, m**	9.01		gong, p	14.85		**goodchild, m**	15.92
2	wang, j	8.52		**goodchild, m**	14.34		gong, p	14.71
3	beven, k	7.14		**openshaw, s**	11.20		**justice, c**	14.13
4	foody, g	6.86		**justice, c**	10.41		**openshaw, s**	13.60
5	curran, p	6.75		wang, j	10.38		estes, j	11.87
6	martin, d	6.49		ledrew, ef	10.37		**li, zl**	11.87
7	stein, a	6.48		shokr, me	9.85		**chen, j**	10.88
8	**openshaw, s**	5.53		ramsay, b	9.71		foody, g	8.79
9	fuller, r	5.23		foody, g	9.34		belward, a	8.48
10	bouma, j	5.11		**li, zl**	9.14		jackson, tj	8.23
13	**chen, j**	4.85	**15**	**chen, j**	8.49	**20**	curran, p	6.85
14	**justice, c**	4.74	**21**	curran, p	6.20	**21**	martin, d	6.76

Table 4.25 Top-Ranking betweenness centrality in the largest components of all networks. Authors in bold hold up their central position almost through all networks.

Furthermore, in contrast to closeness and degree rankings where the presence of lesser-known authors in the GIScience realm is noticeable, the betweenness ranks for all networks show a central group of very familiar authors in the discipline. Moreover, the majority of these central authors in 4-2, apart from J. Kim, SW. Running and GD. Smith, become even more central in the subsequent networks, where topics are closer to GIScience. Moreover, authors such as S. Openshaw, C. Justice, J. Chen, Zl Li and M. Goodchild, from the group of higher betweenness scores (in bold in Table 4.25), are also highly active (Table 4.24). Their collaborations and publications have placed them in favoured positions as they need fewer intermediaries to reach a larger number of other scholars through their large number of collaborators or the collaborators of their collaborators. Likewise, their ideas, findings, or opinions may reach many others quicker than from scholars with an average number of collaborators or low betweens or closeness scores. However, despite their central positions regarding their higher degree, closeness or betweenness scores, one cannot declare to what extent the network topology depends on these authors to be connected. To be able to do so, the next section explores network centralization measures to evaluate to what extent the networks are centralized around a small group of authors with the highest centrality scores.

4.4 GIScience Network Structure

The previous section identified prominent authors, focusing on individual network centralities such as degree, closeness or betweenness. Despite these centrality measures providing insights into the individual's location (ego-centric approach) within the network, they do describe the centralization of the network as a whole. This section focuses on measuring the degree to which the entire GIScience co-authorship networks are focused around a few central authors. In doing so, one can determine whether or not central authors and their connections are crucial to hold the networks together. Three concepts are used for these tasks, centralization measures, component sensitivity analysis, and bi-component identification.

4.4.1 Centralization Measures

Network centralization scores measure the degree to which an entire network is focused around a few central nodes. Thus in a highly centralized network such as a closed terrorist group (Krebs, 2002), the information flow can be controlled by few central nodes. If these nodes are removed, the network is fragmented into isolated sub-networks and information circulation within the network could be interrupted. In the opposite case of a less centralized network, there are no single points of failure, as there are no central nodes on which the overall connectivity or traffic flow depend. Pajek software, which implements Freeman's (1979) centralization measures based on geodesic paths (see section 4.3), was used to calculate the scores. Nooy *et al* (2005) formally defines the measures as follows:

Degree Centralization of a network is the variation in the degrees of the vertices divided by the maximum degree variation which is possible in a *star-network* of the same size. A start network is known to be the most efficient structure given a fixed number of lines (Freeman 1979; De Nooy, *et al*, 2005), offering the highest variations of centralization as all peripheral nodes are connected to the central vertex but not connected among themselves. The degree centralization can vary from 0, which expresses no variation and hence a less centralized network, to 1 expressing the maximum variation in a totally centralized network. In this specific case, degree centralization was computed using all nodes' degree.

Closeness Centralization is the variation in the closeness centrality of vertices divided by the maximum variation in closeness centrality scores possible in a network of the same size.

Betweenness Centralization is the variation in the betweenness centrality of vertices divided by the maximum variation in betweenness centrality scores possible in a *star-network* of the same size. For the betweenness and closeness centralization, only nodes in the largest component were taken into account. As these two centrality measures are based on geodesic distances, then, finding at least one path between all pairs of nodes is a condition. The centralization results, in Table 4.26, show that in general, all networks exhibit a low-level of centralization despite the presence of authors with higher centrality measures (4.3.1, 4.3.2 and 4.3.3). It suggests that the networks are not focused around any particular author.

	Degree Centralization ($x\ 10^3$)	**Closeness Centralization ($x\ 10^1$)**	**Betweenness Centralization ($x\ 10^2$)**
4-2	1.46	1.2636	4.331
4-3	2.33	1.0667	9.612
4-4	3.31	1.0726	11.911
5-2	2.24	1.0549	8.932
5-3	2.58	1.0455	14.694
5-4	3.6	1.209	15.712

Table 4.26 Centralization measures of all networks based on degree, closseness and betweenness centrality. The figures for degree centralization have been multiplied by 10^3, those for closeness centralization multiplied by 10^1 and for betweenneess multiplied by 10^2

Regarding authors' connections, low levels of degree centralization imply relatively less concentrated network structures. This suggests that information and knowledge could be shared effectively through the multiple linkages among the relevant scholars, as the flow in any network does not crucially depend on a few highly active participants. In the specific case of the GIScience networks, at the time of spreading information scholars in networks 5-4 or 4-4 would be far more dependent of those who have more potential communication channels (collaborators) than scholars in networks 4-2 or 5-2. Additionally, rather low closeness centralization measures (see Table 4.26) imply low to medium levels of centralized access to information within the networks. There is not an actor(s) that has a disproportionate access to information compared to others.

However, authors ranked with higher closeness scores in 5-4 or 4-2 networks may find it easier to access information as they can more directly reach many others.

Nevertheless, it is important to note that in the case of 4-2, the closeness centralization result is biased towards the high node degree of scholars in a 68-author paper (see more detail in section 4.2.1). Regarding information flow, there is a highly significant difference in the betweenness centralization scores of 4-2 and the other networks. It shows that, for example, in the largest component of 4-2 there are fewer authors lying on the shortest paths among authors than in 5-4. In other words, it indicates that authors with the highest betweenness centrality scores in 4-2 could exert influence over only 4% of the information flow channels. In contrast, in 5-4, highest between authors could influence over 16% of information flow channels

In summary, smaller networks such as 5-4 and 4-4, whose data sources comprise more relevant journals in GIScience or allied disciplines, appear to be more centralized. Moreover, new connections added by new journals in 5-4 and 4-4 to build 5-3, 5-2 or 4-3, 4-4 networks do not translate into more centralized networks, despite the fact that the number of authors in the largest community do increase (see Table 4.21). These results may be interpreted as a direct consequence of people working around core topics or in a wide range of research themes. People working on similar or related topics are more likely to collaborate between themselves building up a topology that may highlight the presence of central authors in the discipline. On the other hand, an ample network coverage will exhibit a collection of authors who are centrally located in their topics but not in the overall network structure.

4.4.2 Component Sensitivity Analysis

The previous section found that despite the existence of central authors regarding degree, betweenness or closeness scores, the network structures are not totally centralized around them. However, one does not know to what extent the removal of these central authors affects the network connectivity. In that sense, Adamic and Huberman (2002) and Barabasi and Bonabeau (2003) argue that in certain network structures such as the Internet or WWW, the removal of nodes with a disproportionate number of connections will affect negatively the network' connectedness. To test this premise, this research explores the component sensitivity (Moody, 2004) of the GIScience collaboration networks.

In a graph, the component sensitivity gives the size of the largest connected component when all actors with a given degree or more are removed. Therefore, through examining the component sensitivity, one determines to what extent the absence of highly prominent scholars, in terms of number of collaborators, affects the connectedness of the GIScience scientific collaboration networks. Specifically, the component sensitivity analysis was applied to networks 5-2 and 4-2 , as they exhibit the highest and lowest level of degree centralization. The results are depicted in Figure 4-20, where the top panel corresponds to the 4-2 network and bottom panel to the 5-4 network. Each network panel shows the degree distribution (pink and red dots respectively), with the component sensitivity curves in blue. From both networks, the component sensitivity shows that the removal of higher degree nodes, in the fat-tail, does not affect the network connectivity (see blue curve).

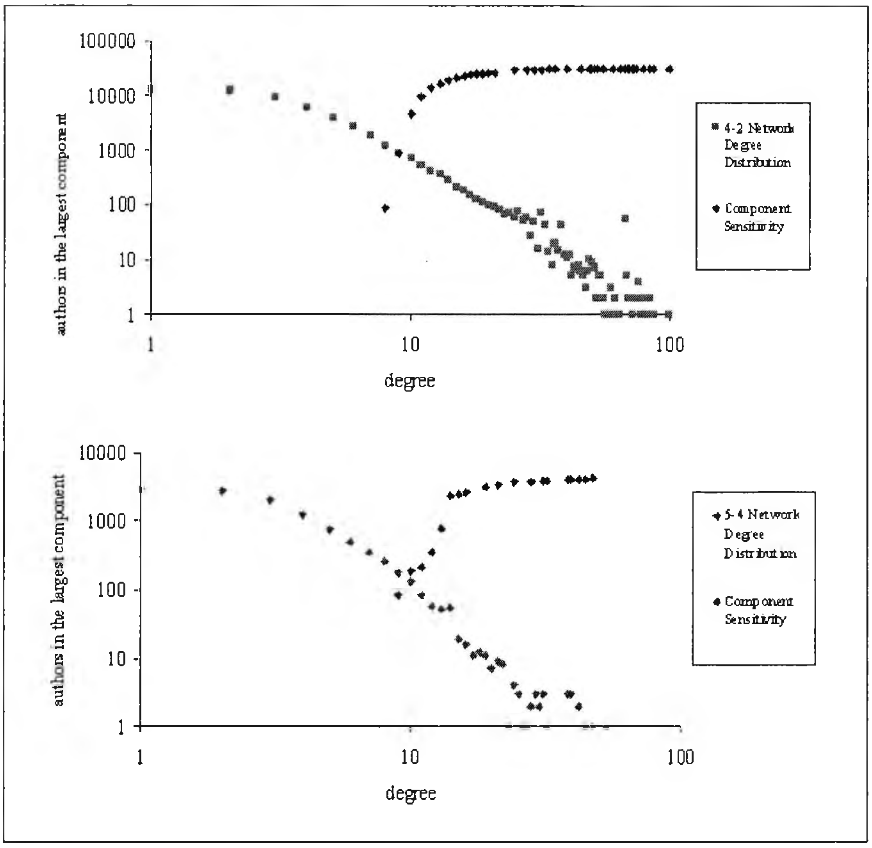


Figure 4-20 Log-Log Plot presenting the component sensitivity analysis of the largest interconnected group in networks 4-2 and 5-4.

Moreover, the networks are unaffected until authors with degrees of around the average of 9 -10 collaborators are removed. In fact, this is equal to removing around 180 and 1000 highly active actors in 5-4 and 4-2, respectively. Hence, as the 4-2 network is less centralized, one needs to remove six times more key authors to fragment it into disconnected sub-networks, than, in the 5-4 network. Consequently, it is clear that neither network is held together by a few authors, a fact that is also reflected in their low scores of degree centralization.

4.4.3 Bridging the GIScience Co-authorship Networks

Results of centralization measures, backed by sensitivity analysis results, revealed fairly decentralized GIScience collaboration network structures. Despite possessing prominent authors, the networks' connectedness does not depend on their connections. As research in GIScience involves multitudes of topics that require the experience and knowledge of scholars from diverse backgrounds, one could think that collaboration networks would form distinct sub-communities. Nevertheless, according to the largest community sizes in Table 4.20, one concludes that as the networks embrace more topics (from the journals), many smaller disconnected sub- communities are linked to the networks' cores. It implies that at some point, there are authors that bridge smaller groups to the network's core. They do so through new co-authorship connections (brought by the added journals) with scholars that are already part of the core. In graph theory, these connections are regarded as *bridges*.

Formally, according to De Nooy *et al* (2005), a *bridge* is a line whose removal increases the number of components in the network. In the same way, a node in a bridge can be regarded as a *cut-vertex,* if its deletion increases the number of components in the network. Hence, one can find bridges from which only one end is a cut-vertex. In this way, cut-vertices are scholars that may act as information brokers, as their privileged positions bring the only access to otherwise disconnected parts of the network. From around 4000 authors in the largest component of the 5-4 network, there are around 600 (the number of bridges were calculated by determining the number of bi-component of size 2 with Pajek) authors who are parts of bridges. From approximately 32000 authors in the largest component of 4-2, around 4800 scholars form links that bridge smaller groups to the core network.

These figures show that within both networks around 15% of the authors play pivotal roles within the network structures. Therefore, the removal of these authors and their links, rather than the ones with the highest centrality scores, will severely affect the network connectedness. In addition, one can argue that the larger the size of the component without cut-vertexes, the less reliant on them the network structure becomes. As previously mentioned in section 2.2.2, the largest or one-component is a fragile structure since, as Moody (2004) explains, the removal of a single person could disconnect the network. In contrast, a bi-component is a relatively less vulnerable connected sub-network, of minimum size 3 that does not contain a cut-vertex (De Nooy *et al*, 2005). In that sense, a bridge can be taken as a bi-component of size two without a cut-vertex (De Nooy *et al*, 2005). Furthermore, one can say that a bi-component is a more cohesive collaboration sub-network (Moody, 2004; De Nooy *et al*, 2005), because there has to be at lest two different co-authorship paths between any pair of authors.

Bi-Components Statistics	**no authors**	
	5-4	**4-2**
In the largest or one-component	4227	31886
In the largest bi-component (taking into account authors that belong to more than one bi-component)	1341	15212
Belonging to more than one bi-component or bridge (key authors)	799	5802
number of bridges (bi-component with size 2)	637	4889

Table 4.27 Bi-Component and bridge statistics in networks 5-4 and 4-2

According to Table 4.27, around 30% of authors in 5-4 and 47% in 4-2 largest component are also members of the largest bi-component. The figures indicate a high proportion of authors very likely to belong to similar research specialities. In total, there are approximately 600 bridges in 5-4 and 4900 in 4-2. Moreover, around 18% of the authors belong to more than one bi-component or bridge. This is the group of authors that in fact, plays a key role maintaining the GIScience networks' connectedness.

Table 4.28 shows the group of pivotal authors who are part of the bridges or bi-components. It is interesting to note in the case of the 4-2 network that, even though the listed authors hold strategic positions bridging different parts of the network, they do exhibit relatively low centrality scores. Table 4.28 shows that within ranked authors for 4-2, there is only one in the 100th top degree list and six in the 100th betweenness ranking. Moreover, only two of the listed authors (lee, j. and king, d.) in the table have high node degree and betweenness

Author	No Bridges or No Bi-Components that author belongs to	Degree Rank	Closeness Rank	Betweenness Rank
butler, j	17	1047	22843	1164
grant, w	13	178	10298	295
lee, j	12	93	363	35
nijkamp, p	12	205	25560	767
pedrycz, w	11	769	14581	975
singh, vp	11	152	1997	57
oppenheim, c	10	1638	4527	512
yan, h	10	1898	6928	1682
kamel, m	9	1791	6771	2485
kandel, a	9	299	2040	123
lambin, ef	9	994	402	397
recknagel, f	9	1015	17479	1047
stein, a	9	104	249	66
bishop, i	8	803	4039	500
brown, rd	8	1544	19873	2546
burn, dh	8	2591	11707	2629
buttle, j	8	952	1036	530
davis, re	8	736	1347	1112
fraser, cs	8	3572	18186	1187
gupta, a	8	324	2603	148
hall, r	8	1065	5197	610
hancock, er	8	2680	29620	2406
harris, r	8	4326	18375	3046
king, d	8	118	205	58
kittler, j	8	1432	31359	563
kong, l	8	3693	30190	2181
mckee, m	8	145	4988	231
meeker, wq	8	3778	6726	2976
rees, wg	8	773	3257	811
svirezhev, y	8	3938	3307	2654
tanaka, k	8	1869	15292	1690
taylor, j	8	478	16310	706
yang, gcc	8	2527	6022	1020

Table 4.28 Network 4-2's strategic authors (those who are part of the highest number of bi-components or bridges) and their centrality scores.

A very different case is presented by the 5-4 network, where the majority of key players listed in Table 4.29, also have top degree and betweenness scores; however, not many are among the top closeness centrality. In both networks, this fact shows that bridging authors are more likely to have many connections and along shortest paths between others, rather than being close to many of them

Authors	No Bridges or No Bi-Com that author belong to	Degree Score Rank	Closeness Score Rank	Betweennes s Score Rank	Authors	No Bridges or No Bi-Com that author belong to	Degree Score Rank	Closeness Score Rank	Betweenness Score Rank
egenhofer, mj	12	12	189	30	fotheringham, as	5	177	176	158
goodchild, m	9	6	6	1	franklin, se	5	30	1175	274
congalton, rg	8	65	632	127	hassan, ma	5	440	2667	77
karnieli, a	8	10	68	37	jezek, kc	5	183	2471	174
rees, wg	8	45	1531	148	khorram, s	5	282	544	391
cracknell, ap	7	2	173	15	king, d	5	447	3014	516
lambin, ef	7	138	86	388	kraak, mj	5	450	1187	329
robinson, is	7	50	1171	51	lee, j	5	460	3708	199
singh, rp	7	83	834	205	li, x	5	229	260	258
allison, rj	6	517	1808	476	li, zl	5	23	10	6
armstrong, mp	6	391	2708	429	macklin, mg	5	347	2709	108
carr, jr	6	259	2026	442	martin, d	5	42	129	21
church, m	6	717	3244	349	mason, dc	5	34	2350	324
curran, p	6	3	101	20	mather, p	5	816	4021	278
frank, au	6	217	877	204	parsons, aj	5	357	2258	468
jensen, jr	6	5	34	23	price, kp	5	26	134	45
lo, cp	6	465	1189	371	raper, j	5	1135	2298	69
timmermans, h	6	63	2759	292	rigby, d	5	629	981	444
toutin, t	6	158	3408	276	roberts, a	5	362	2498	207
anderson, jm	5	387	1547	142	saraf, ak	5	21	2441	85
barber, dg	5	316	1118	477	setzer, aw	5	1509	1721	632
burrough, pa	5	533	1900	443	stuart, n	5	658	2987	76
campbell, h	5	535	3001	328	thorne, c	5	510	4105	392
cavayas, f	5	404	2385	348	woodcock, c	5	16	178	87
dubois, jmm	5	264	1389	104	wooster, mj	5	250	1741	327
faig, w	5	1315	3367	631					

Table 4.29 Network 5-4's strategic authors (those who are part of the highest number of bi-components or bridges) and their centrality scores.

Nevertheless, there is a significant difference between the listed authors in 4-2 and 5-4 networks. While in network 4-2, the list does not yield many known authors in the GIScience domain, 5-4's list seems to yield within the research community not only well-known scholars but, scholars with very different research interests. The topics from these authors range from visualization, remote sensing, core topics, etc. However, further analysis needs to be done regarding this point.

This chapter sheds some light on patterns of collaboration and co-authorship extracted from six GIScience networks with different coverage of topics in the area. They show that on average papers are written by teams of around two authors. Authors in the network are participating with less than two papers and collaborating on average with five others. These patterns lead to a largest connected community of around 50% of the authors in the most general network, but not larger than 30% for the smaller network that covers topics very relevant to GIScience. Within the largest communities, around 15% forms more cohesive groups linked by at least two independent co-authorship paths.

Additionally, centrality measures show the presence of few scholars with a disproportionate number of collaborators, who are close to many others and in between the most efficient paths linking many pairs of scholars. Nevertheless, the centralization measures show that despite the presence of these "star authors", the networks' structure is not focused around them. The result of the component sensitivity analysis shows that one needs to remove more than 500 of the top degree authors in order to seriously fragment the network. In contrast, there is a group of authors, who are not central regarding degree, betweenness or closeness, but belong to many bi-components or act as bridges between parts of the network, otherwise disconnected. These are the authors who are truly holding the network structure together. One can speculate, especially in the 5-4 network, that they may be authors working in various research areas that cross collaborate with many others, resulting in links that join different research specialities. Trying to corroborate this assertion, the next chapter will attempt to link the revealed network structure to GIScience practice in order to establish to what extent this is a multidisciplinary domain and the listed authors are representative of each one of the related sub-specialities.

5 Multidisciplinarity and GIScience

The previous chapter examined the basic statistics and topological properties of six GIScience co-authorship networks built from six different, but nested, bibliographic data sets. Each network is believed to represent a particular instance of scientific collaborations around topics with different levels of relatedness to GIScience. The variety of results obtained from each network shows how the inclusion or exclusion of journals (that represent research topics) affects the positions of the authors within the network topology. Moreover, as journals represent research topics, each network exhibits a distinct structure that is the product of multidisciplinary and inter-disciplinary scientific interconnections around basic or applied GIScience research. Thus, each network represents a different inside look at the multidisciplinary nature of the domain.

Due to the nature of the journals involved in the 5-4 network, its network topology represents scientific collaborations around key GIScience topics and issues in allied disciplines. Therefore, one can consider that 5-4 outlines the core of all GIScience collaboration networks under study. Hence, this chapter adopts network 5-4 as an instrument for the analysis of the multidisciplinary nature of scientific collaborations revealed by co-authorship links extracted from core and allied journals to the discipline.

To this point, co-authorship networks have been studied as a whole, aggregating journals into single data sources. In order to analyze the multidisciplinary nature of the GIScience core representation, one needs to break down information at journal level. As a result, this chapter aims at analyzing the extent to which co-authorship contributions from each research speciality (representing by journals) helps in shaping the core's structure. The chapter is divided into three sections. The first section outlines the general features of each journal and identifies its research domain and relationship to GIScience. The second explores patterns of collaboration and co-authorship in order to find out the contribution of each journal in shaping the overall network's patterns. The final section introduces statistical entropy as an instrument to quantify the degree of multidisciplinarity.

5.1 A View at Journal Level

The publication outlets of the core network comprise the 10 closest peripheral journals to IJGIS (see Table 5.30). They cover basic research and computing aspects related to GIScience, such as in IJGIS, Geoinformatica or selected LNCS.

Journal Name	Acronym	Research Focus (Goodchild, *et al*, 1999)	Research Stream (Mark, 1999)
Computers & Geosciences	C&G	System	Basic
Computers Environment And Urban Systems	CEUS	System, Society	Applied
Earth Surface Processes And Landforms	ESPL	Society	Applied Allied Discipline
Environment And Planning A	EPA	Society, Individual	Applied
Environment And Planning B-Planning & Design	EPB	Society, Individual	Applied
Geoinformatica	Geoinformatica	Systems	Basic Allied Discipline
International Journal Of Geographical Information Science	IJGIS	Systems Society Individual	Basic and Applied
International Journal Of Remote Sensing	IJRS	System Society	Applied Allied Discipline
Lecture Notes in Computing Science	LNCS	Systems	Basic Allied Discipline
Photogrammetric Engineering And Remote Sensing	PERS	Systems	Applied Allied Discipline
Progress in Human Geography	PHG	Society Individual	Applied Allied Discipline

Table 5.30 Research Focuses and Research Streams of the Core Journals within the GIScience Collaboration Network. The acronyms will be used to identify each journal from now onwards. IJGIS, the seed journal, is highlighted in light grey.

The publication outlets also include journals covering developments from applied research from various allied disciplines such as C&G, ESPL, IJRS, PERS and PHG, CEUS, EPA or EPB. The inclusion of these journals reflects importance of applied research within GIScience (see section 3.1.2 chapter 3). As representing scientific interactions between authors in the core and peripheral journals, they are regarded as "*GIScience core*" within the context of this chapter's delivery and discussions. The following sub-sections discuss the main GIScience research questions and their relationship with the selected journals. Additional, journal publication patterns and their contributions to the structure of the overall GIScience network are presented

5.1.1 GIScience Publication Forums

Within scientific communities, academic journals are regarded as a formal system for communicating task outcomes within the scientific communication process (von Unbegrn-Sternberg, 2000; Whitley, 2000; Borgman and Furner, 2002; Rey-Rocha and Martin-Sempere, 2004). As von Unbegrn-Sternberg (2000) emphasizes, every major traditional field has a few high status and influential journals where content is controlled by a small set of gatekeepers, and is widely read within its scholarly community. Thus, scholars seek to publish there as a way of communicating their research results to the correct audience, and as Whitley (2000) states, to acquire reputation within their own disciplines. In that way, to get published implies that fellow researchers are convinced of the importance and significance of the results (Whitley, 2000; Borgman and Furner, 2002). Likewise, as authors' own views and ideas are accepted as important, others are more likely to follow their directions, helping them establish higher positions and reputations within their academic communities (Whitley, 2000). Traditional disciplines such as modern physics or chemistry possess a well-defined set of journals considered as "high-quality", where only novel and cutting edge research is published (Whitley, 2000). Borgman and Furner (2002) note that such as journals attract a very large number of manuscripts entailing a very low rate of acceptance and conversely very large rejection rate.

In contrast, GIScience observes a different panorama, where there is not an exhaustive list of journals (Arciniegas and Wood, 2006) that should be considered as universal publishing space for scholars working on topics relevant to the community. The major factor accounting for that may be related to intrinsic characteristics of the research on GIScience per se. In general terms, research on GIScience, according to Mark (1999), can be divided into two very deeply interconnected streams, one that addresses basic GIScience research (see section 3.1.2. chapter), and the other concerning research using GITechnologies such as GPS, GISystems, remote sensing, etc. The former, basic research, requires the expertise of scholars from diverse backgrounds such as geographers, statisticians, mathematicians or computer scientists (Goodchild *et al*, 1999). On the other hand, research applying GITechnologies requires scholars with a wider range of expertise due to the diversity of problems in which *space* plays an important role (Goodchild, 1992; Wright *et al*, 1997; Mark, 1999; Mark, 2003). Hence, the result is a discipline in which scientific problems are surely best addressed through multidisciplinary collaboration rather than by the expertise of scholars from only one discipline or domain (Goodchild *et al*, 1999).

Goodchild *et al* (1999) and Cova (2000)[7] group the crucial issues needing to be addressed by GIScience in three distinct arenas. One centres around *individuals* and their role as users of GI and GITechnologies. The second group, the *system* focuses on all issues surrounding GITechnologies such as hardware, software, etc., and the last group focuses on all issues encompassing the *society* and how it is affected by the use of GI and GITechnologies. As a result of using this classification, Goodchild *et al* (1999) identify areas where GIScience may benefit from the experience and expertise of their members. Thus, one can list some disciplines that are most likely to contribute to each of the research streams. For example, Goodchild *et al* (1999) proposed for the individual group: cognitive science, environmental, psychology or linguistics, for the system: computer science and information science, and for society: economics, sociology, social psychology, geography and political science.

Therefore, scientific progress in GIScience is expected to be covered not only by journals focused on traditional GI issues, but also, by journals that cover progress in allied disciplines. In that way, Table 5.30 shows the core participant journals classified according to which research stream they relate to (basic or applied), and to which areas they are focused on (individual, society or systems). The list includes journals of various kinds such as IJGIS and C&G considered core journals in the discipline, from geosciences such as ESPL or PHG, and also, from disciplines where GI and GITechnologies play an important role such as IJRS, PERS, EPA or CEUS.

5.1.2 Patterns of Published Scientific Documents on GIScience Core Network

As previously explained, a co-authorship network is the result of connecting authors via collaboration links from published scientific documents in academic journals. Frequently, studies in co-authorship networks neglect other types of documents rather than articles (Moody, 2004). Despite that, this study takes into account all types of published scientific documents that ISI-WoK has indexed from the core journals. In doing so, typical document types from the research topics or disciplines represented by the journals under study are examined and compared.

Unsurprisingly, Table 5.31 shows that the most frequent type of publications within the core corresponds to academic research articles or papers, which represents almost the 70% of the documents. Surprisingly, the second most frequent type corresponds to 20% book reviews, followed by around 6% of editorial material.

[7] Cova (2000) reports also on a fourth issue concerning the Earth.

Type of Scientific Document	No in the Network	%
Article	8413	68.00%
Book Review	2487	20.08%
Editorial Material	718	5.80%
Review	204	1.60%
Note	202	1.60%
Letter	104	0.80%
Item About an Individual	55	0.40%
Software Review	45	0.40%
Correction, Addition	43	0.30%
News Item	39	0.30%
Correction	32	0.30%
Biographical-Item	22	0.20%
Reprint	6	0.00%
Bibliography	2	0.00%
Database Review	1	0.00%
Total Documents in the Network	**12373**	

Table 5.31 Type of documents contributing to the GIScience core network

The representation of other less common types such as review, note, letter, software review, or reprint corresponds to around 6%. Overall, it seems that books are an important part of GIScience, as reviews about them are printed along side papers in the journals. Therefore, the lack of book or book chapter coverage by WoK neglects an important part of GIScience scientific collaboration and so affects the exhaustiveness of this study.

Interestingly, individual journals show considerable variation in preferences in terms of document types. For a better insight, the 11 journals are divided into 4 categories regarding the nature of the topics covered (see Table 5.30). This journal classification will be used within the context of this chapter. The first category represents journals considered as core for the discipline such as IJGIS, C&G, LNCS and Geoinformatica. Regarding this group, Figure 5-21 shows that articles are the most frequent type within this group, but book reviews play an important role in IJGIS. Roughly, IJGIS and C&G follow the pattern of the overall network. This pattern varies from the remaining two journals, LNCS and Geoinformatica, as they only cover articles.

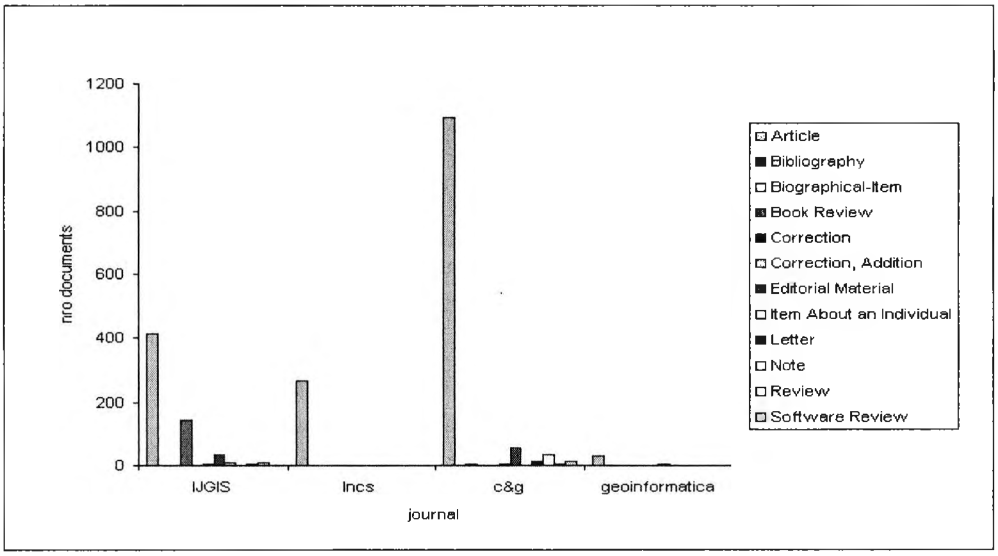


Figure 5-21 Document type distribution. Group One - Core GIScience and Computer Science Journals.

The second group corresponds to journals focused on planning and environmental issues such as EPA, EPB and CEUS. Among them, Figure 5-22 shows that article are the preferred type. However in EPA and EPB, book reviews are more frequent, to the point of that in EPB the difference between articles and book reviews is very small. Thus, the importance of books within the environmental research community it is also evident.

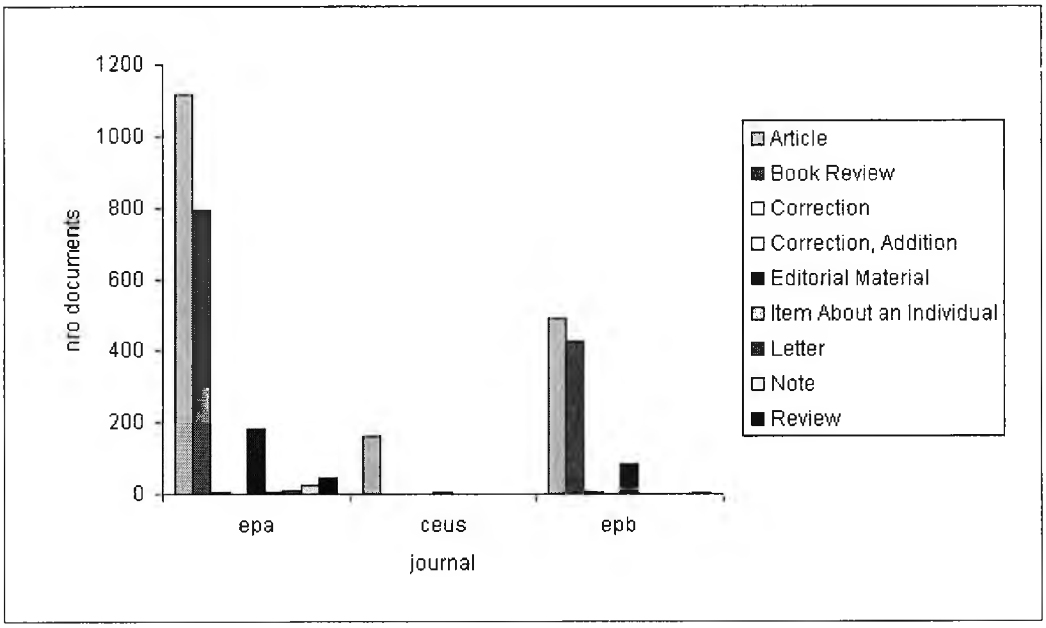


Figure 5-22 Document type distribution. Group Two - Planning and Environmental Journals.

The third group comprises PERS and IJRS, journals publishing work on remote sensing issues. According to Figure 5-23, papers prevail as scientific communication documents within the two remote sensing journals. In contrast to IJGIS and environmental journals, the presence of books reviews or editorial materials are almost imperceptible in this group.

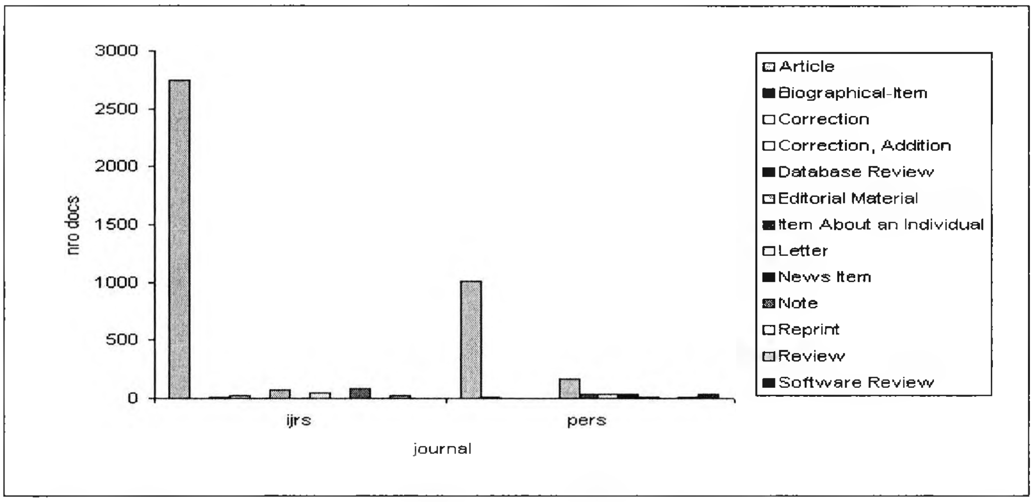


Figure 5-23 Document type distribution. Group Three - Remote Sensing Journals

Finally, the fourth category groups PHG and ESPL, journals from the geoscience arena. Figure 5-24 reveals that the two journals regarding type of published documents follow completely different patterns. In PHG the most frequent document type by far is the book review, rather than article like in the previous groups. It reveals the importance of academic books within the human geography subfield. In ESPL the article is the preferred type of document, but more important than book review s are editorial material.

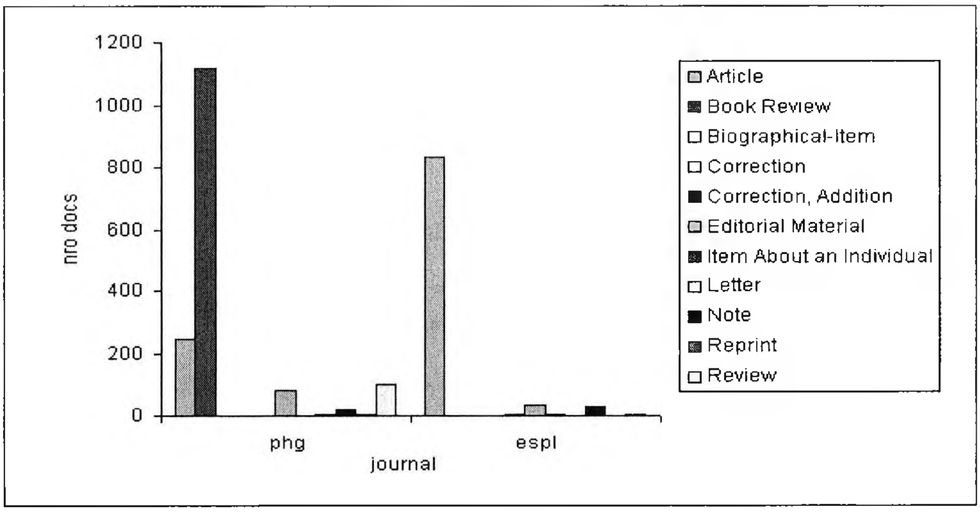


Figure 5-24 Type of document distribution. Group Four - Geosciences Journals

The previous two sections presented the core journals regarding their research coverage and focuses, and differences in the type of published academic documents. The next section evaluates each journal contribution to the core network topology in terms of number of papers, in order to find out which are the predominant journals.

5.1.3 Journals in the GIScience Co-Authorship Network

There is not a uniform contribution of documents from each journal to the core network. As Table 5.32 shows, there are some journals such as IJRS, EPA or PHG with more papers than the others. The difference between this network totals and those in Chapter 4 corresponds to 80 papers excluded from Geoinformatica. In fact, they are manuscripts from *Transactions in GIScience* (see explanation in Chapter 3).

Journal	Frequency (No Issues per Year)	Journal Contribution (in No. papers)	
IJRS	12	3015	24.4%
EPA	12	2180	17.6%
PHG	4 up to 2001 6 in 2002	1572	12.7%
PERS	12	1355	11.0%
C&G	10	1223	9.9%
EPB	6	1019	8.2%
ESPL	13	914	7.4%
IJGIS	10	626	5.1%
LNCS	na	266	2.1%
CEUS	6	169	1.4%
GEOINFO RMATICA	4	34	0.3%
Network Total		**12373**	

Table 5.32 Journal share of the total number of documents in the GIScience co-authorship network

Intrinsic and extrinsic factors to the journal may account for the number of papers per journal indexed by the ISI-WoK. The former involves factors such as the number of issues or number of papers per issues; figures controlled and defined by the publisher. Accordingly, the number of issues per year varies from journal to from quarterly, half-yearly, monthly, etc (see Table 5.32). Moreover, within the same journal the issue number is also variable. As in the case of PHG that published 4 issues between 1992 and 2001 and increased this number to 6 from 2002 onwards. Besides, there are occasional special issues that cover a specific leading edge topic or best papers in a domain-related conference. Additionally, there are extrinsic factors that are not controlled by the journals editorial board such as time coverage in the bibliographic database. Fisher (2001) argues that the reason for the absence of important core GIScience journals such as *Transactions in GIS*, *Geoinformatica*, *Geographical Systems* or *URISA* in the ISI-WoK database, was because in that year the journals were too recent to be included. However, there were still not indexed in 2005, when the data source for this project was downloaded and assembled.

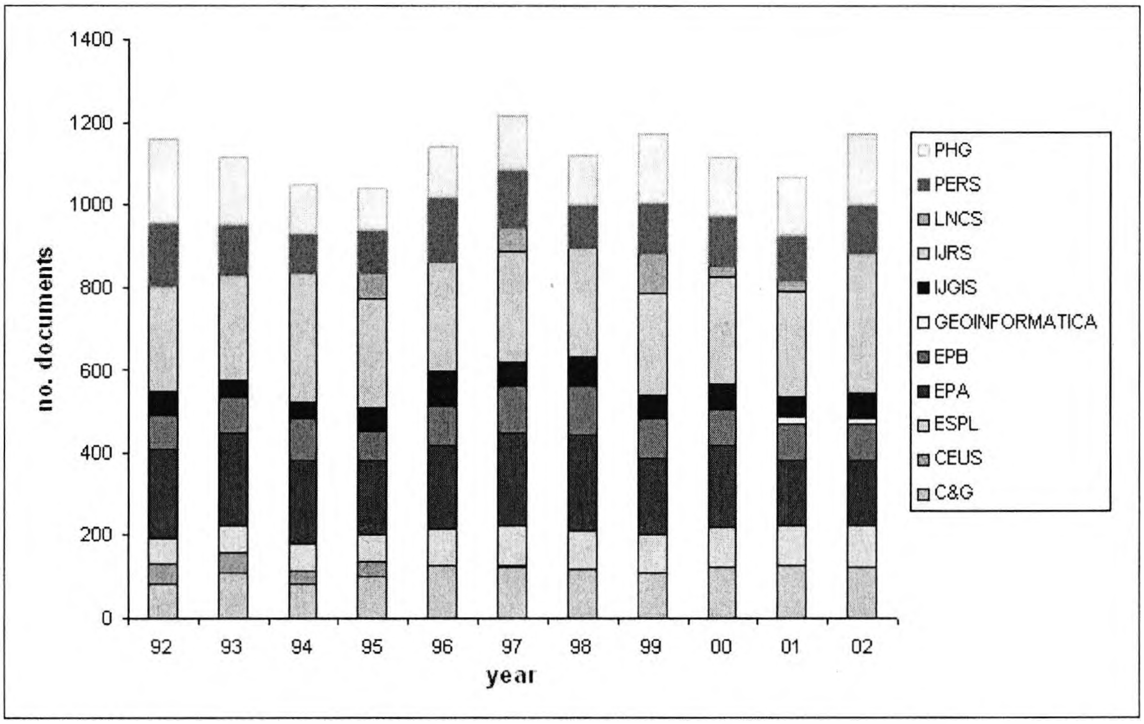


Figure 5-25 Journal contribution in number of documents to the co-authorship network between 1992 and 2002. Yellow shade corresponds to Geosciences, pink shade to Remote Sensing, blue shades to core GIScience and Computer Science. Green shades represent the group of journals from Environmental issues. Each category is consistently represented by the same shades in all charts throughout this chapter.

Both, intrinsic and extrinsic factors affect the total number of documents per year that each journal contributes to the co-authorship network. Figure 5-25 depicts the percentage of papers by journal contributing to the network for the time period. In the case of Geoinformatica although it was first published in 1997, ISI-WoK only indexed it from 2001. Similarly, CEUS is indexed only between 1992 and 1995 and LNCS has not a regular contribution to the network as only certain issues related to the GIScience topic were selected (see Chapter 3).

Consequently, each journal helps to shape the GIScience network topology differently according to the number of documents included. Figures in Table 5.32 illustrate that journals such as IJRS, EPA and PHG contribute with more than 50% of the papers to the GIScience network. In contrast, documents from journals such as Geoinformatica, CEUS or LNCS represent less than 4% of the total. Accordingly, central authors in the highest contribution journals' local networks are more likely to have a stronger influence on the overall network topology of the core. Nevertheless, one cannot say to what extent, as this depends on the specific patterns of collaboration and authorship of the authors per se and their collaborators. Hence, the following sections 5.2 and 5.3

concentrate on to examining the position of each journal and their contributors within the overall network topology.

5.2 Core GIScience Collaboration and Co-authorship Patterns

Individual journals cannot be used to generalize patterns within the disciplines or fields they represent and cover (Börner *et al*, 2004; Moody, 2004). However, the patterns of individual journals may give an insight that enables their role to be discerned in shaping the overall network structure. In doing so, one can evaluate whether or not the predominant journals (based on their individual measures) may exert stronger influence than the others over the overall network topology.

Network	No. Authors	No. Papers	Network Participation (mean papers per author)	Co-authorship (average authors per paper)	% Sole Papers	Average no. of collaborator s per author	% Authors connected in the largest component
5 - 4	**13815**	**12373**	**1.75**	**1.96**	**52%**	**4.79**	**30.60**
C&G	1830	1223	1.31	1.96	43%	1.92	16.23
CEUS	256	169	1.16	1.75	49%	1.37	2.34
ESPL	1395	914	1.51	2.30	30%	2.82	21.43
EPA	1890	2180	1.67	1.45	72%	1.30	3.65
EPB	919	1019	1.59	1.43	22%	1.41	2.50
GEOINFO RMATICA	69	34	1.01	2.06	29%	1.39	5.80
IJGIS	892	626	1.36	1.93	48%	2.31	10.76
IJRS	5030	3015	1.64	2.73	21%	3.71	37.85
LNCS	471	266	1.31	2.33	29%	2.44	9.98
PERS	2151	1355	1.39	2.21	42%	3.00	9.58
PHG	860	1572	1.93	1.06	96%	0.36	1.16

Table 5.33 Patterns of co-authorship and collaboration in the core at journal level

This section focuses on analyzing basic network properties, computed in the same manner as figures calculated in chapter 4 (see Table 4.1 and 4.5). For the overall core network (5-4 row in Table 5.33), authors participate on average with less than 1.7 papers. Breaking it down by journal, the average author participation ranges from 1.0 to 2.0 papers (see Table 5.33). In the case of journals with averages close to 1.0 such as Geoinformatica or CEUS, it may be due to the poor bibliographic coverage in ISI-WoK. In the case of journals with an average close to 2.0 such as PHG, EPA , IJRS, EPB or ESPL, results show that authors are more likely to publish twice during the study period. Besides, Table 5.33 reveals that PHG exhibits a greater author participation average than the overall 5-4 network. This result suggests that very participative authors in PHG do not publish papers in the other journal in the network. In contrast, lower averages than the overall network may indicate that some authors from these journals publish in more than one journal within the network.

The co-authorship figures indicate that in 7 journals out of the 11, papers are written by teams of around 2 authors. The largest average co-authorship teams are exhibited by IJRS, in contrast to PHG, where papers are mostly sole authored. Probably, the large amount of book reviews in PHG may account for the 96% of single-co-authored papers (see section 5.1.1) as book reviews are mainly individual contributions. However, in other journals a high percentage of sole-authored papers does not mean low co-authorship rates. For example, 42% of papers in PERS are sole author works, but, the collaborative papers are written by teams of two authors on average. Conversely, a high percentage of collaborative papers does not imply larger co-author teams. For example, EPB has 78% of joint papers, but the teams are on average around 1.50 authors. Nevertheless, there are also journals with a high percentage of collaborative papers such as ESPL, Geoinformatica or LNCS, which on average are written by teams of more than two authors. Finally, there is a group that includes CEUS, IJGIS and C&G, where around half of the papers are joint works written by teams of around 1.8 authors; co-authorship average similar to the overall 5-4 network.

As a result of highly collaborative papers, the two remote sensing journals, IJRS and PERS, have a high average of almost 4.0 and 3.0 collaborators respectively. Interestingly, it seems that higher collaboration levels such as exhibited by PERS, IJRS, ESPL, LNCS and IJGIS lead to well connected authors. However, it is not always the case, as in Geoinformatica where despite its high co-authorship average, authors have around one co-authorship connection on average.

The previous patterns of collaboration and co-authorship at journal level will affect the way that authors connect to one another, forming communities via co-authorship links of different sizes. Hence, it is not surprising that a journal such as IJRS has the highest level of connectedness of around 40% of the authors linked. However, what is more surprising is that this level is larger than the overall 5-4 connectedness level of around 30%, showing a very collaborative and interconnected remote sensing community represented by this journal. One can say that the IJRS interconnected sub-community is embedded in the largest core of network 5-4, as there is no significant sub-communities outside the largest component (see table 4.9 chapter). Besides, journals such as EPA or C&G despite their low co-authorship and collaboration averages, exhibit around 16-20% of interconnected authors. This may show communities where authors do not like to collaborate in larger teams, but do like to collaborate among themselves. At the other extreme, there are journals such as ESPL, where despite its high collaboration and co-authorship averages, less than 3% of the authors are interconnected. It shows very collaborative authors that like working in large teams in this journal, but who do not tend to have diverse collaborators.

This section reveals a group of journals that exhibit very distinctive patterns of publication, co-authorship and collaboration. Though, individual journals cannot be used to generalize patterns within the disciplines and fields they represent and cover. Instead, their individual patterns give an insight that enables each journal participation within the overall network structure to be discerned. However, the section exposes that 3 journals account for more than 50% of the papers. It means that central authors (regarding degree, betweenness or closeness) writing on the specific topic(s) from these journals are most likely to play a central role within the overall network topology. However, the centrality of predominant journals is determined by the patterns of collaboration and co-authorship. A group of central authors regarding degree, betweenness and closeness were identified in Section 4.3. However, it is still unknown in which topics (journals) their work is published. Do they diversify their publications across all 11 journals to exhibit truly multidisciplinary research interest? or Do they concentrate their work on specific research interests? The following section identifies to what extent authors concentrate or spread their publication. In doing so, one will be able to discern the truly multidisciplinary nature of the network regarding co-authorship links around the central topics to GIScience.

5.3 The Multidisciplinary Nature of the GIScience Network Core

The fuzzy research boundaries between GIScience and its allied disciplines (Arciniegas and Wood, 2006) are the result of multiple collaborations between scholars from multiple backgrounds that attempt to solve complex research questions varying in nature, but with a spatial component as commonality (Wright *et al*, 1997; Goodchild *et al*, 1999; Mark, 2003). This multidisciplinary nature is reflected in the variety of topics from the journals selected to build the GIScience collaboration core. This section aims to explore this multidisciplinary nature by analysing publication preferences from the participating authors. In doing so, one can analyze to which degree authors diversify their publication sources, which will indirectly reveal their research interests.

5.3.1 Multidisciplinary Authorship

In building the co-authorship network each journal contributes a particular number of academic documents (see Table 5.32) that are co-authored by different numbers of authors. The differences in co-authorship and collaboration analyzed in section 5.2, reveal very specific patterns of doing research on topics related to the journals under study. However, global patterns do not allow observing the contributions of authors from each journal to the overall network structure. If possible, it would allow studying the degree to which authorship from each participant research topic contributes to the network structure. In that way, the role of each journal and their authors in shaping the network structure could be assessed. This study assumes a direct relation between topic(s) covered by journals and the research interests of the authors contributing to them. Hence, if an author publishes a document in any of the journals included, one assumes that to a certain extent, this author is academically related to the research area covered by the journal.

Table 5.34 shows the total number of authors contributing to each individual journal. It is important to note that authors may publish in various journals; therefore, the total number of authors (13815) in the network differs from the total number of authors publishing papers (24202). The latter was used to compute the following percentages.

Journal	Authorship Share per Journal (in number of authors)
IJRS	34.0%
EPA	13.0%
PERS	12.4%
C&G	9.9%
ESPL	8.7%
PHG	6.9%
EPB	6.0%
IJGIS	5.0%
LNCS	2.6%
CEUS	1.2%
GEOINFORMATICA	0.3%

Table 5.34 Each journal authors share of the total number of authors in the GIScience network.

The figures in Table 5.34 show that almost 35% of the authors in the co-authorship links are from papers published in IJRS. Moreover, as PERS also focuses on remote sensing, the share of authors working on remote sensing represents 46%. Moreover, it reveals that only 19% of scholars' contributions are from core GIScience works (IJGIS, C&G and LNCS). Besides, while 20% of the authors have works on environmental issues covered by EPA, EPB or CEUS, 16% have participated in research that involves Geosciences' topics concerned by PHG or ESPL. The overall results in the table confirm the strong presence of scholars, who at any point during 1992 and 2002 have worked on remote sensing issues. As observed in previous sections, there are not only more in number but also, their particular patterns of collaboration and co-authorship (see sections 4.3 and 4.4, chapter 4) have given many of them central places within the core GIScience network.

Until this point, results have only concentrated on general features that do not allow evaluation of authorship patterns at journal level (research interests from authors with highest centrality measure manually looked for). Consequently, this section concentrates on analysing all scholars' academic contributions to the network, broken down at journal level. Thus, one will be able to evaluate to what degree the academic contribution of the participating authors are well spread among the journals or concentrated in few journals. If not concentrated, this would indicate that authors work on a variety of research topics, exhibiting multidisciplinary activity. Otherwise, the network structure would be shaped by authors working on more specific topics. Therefore, to analyze the multidisciplinary nature of authorship (represented by network participation) in the core of GIScience, this section aims to explore publication patterns exhibited by the most participative authors. One expects to reveal their primary research focuses represented by the journal topics they concentrated their publications on, and to identify to which degree their participation is oriented towards individual or multidisciplinary topics.

Section 4.1.2 in the previous chapter, identified a group of authors with levels of network participation higher than the network average. However, the degree of contribution to the network topology varies according to the collaborative and multidisciplinary nature of their academic work. Figure 5-26 reveals that R. Johnston, the most participative author, does so mainly through individual works (dotted pattern). Similar patterns are observed by J. Butler, P. Longley, M. Batty, Gl. Clark and P. Fisher, for whom more then 60% of their participation is from individual works. Moreover, as a result of contributing only through individual papers (and not having co-authorship links), Ca, Staab and N. Castree are not part of the network.

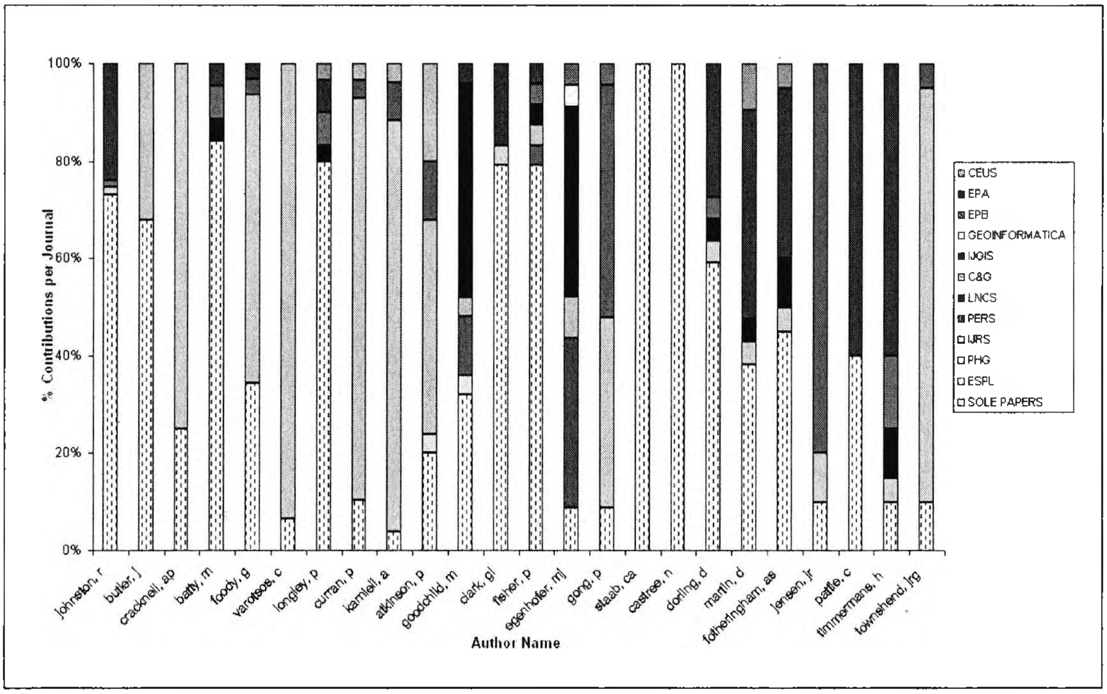


Figure 5-26 Journal distribution of the top-20 participative (sole and collaborative works) authors in the network. Dotted pattern corresponds to sole works, while the shades correspond to the palette for the fourth categories established in Figure 5-25. The figure shows top 24 authors as Fotheringham, Jensen, Pattie, Timmermans and Townshend , the top less participative authors, (on the right hand side) have the same network participation figure (20 papers).

Authors such as AP. Cracknell, C. Varotsos, P Curran, P. Atikson, MJ. Egenhofer, P. Gong and D. Martin observe the opposite behaviour with more than 60% of joint contributions. It is clear that some authors achieve their top positions through participating mainly with sole papers (see 1st, 2nd, 4th, 7th, 12th, 16th -18th).

As a co-authorship network is built upon paper resulting from scientific collaborations, if one excludes sole works, Figure 5-27 show a similar top author list though in different order. Here, top authors are those who contribute with the higher number of collaborative works to the network. Authors such as R. Johnston and R. Butler are included despite participating with a high number of sole papers, as their collaborative network participation is higher than the network's average. Additionally, Figure 5-27 points out the multidisciplinary nature of the collaborative contributions from some of the top authors. As different colour shades indicate the four different research topic categories (see Figure 5-25), the strong influence of remote sensing (pink shades) from the top authors works is evident. It indicates that authors such as AP. Cracknell, C. Varotsos, P. Gong, JR. Jensen or JRG. Townshend contributes with a high number of joint works, but only do so, on remote sensing topics.

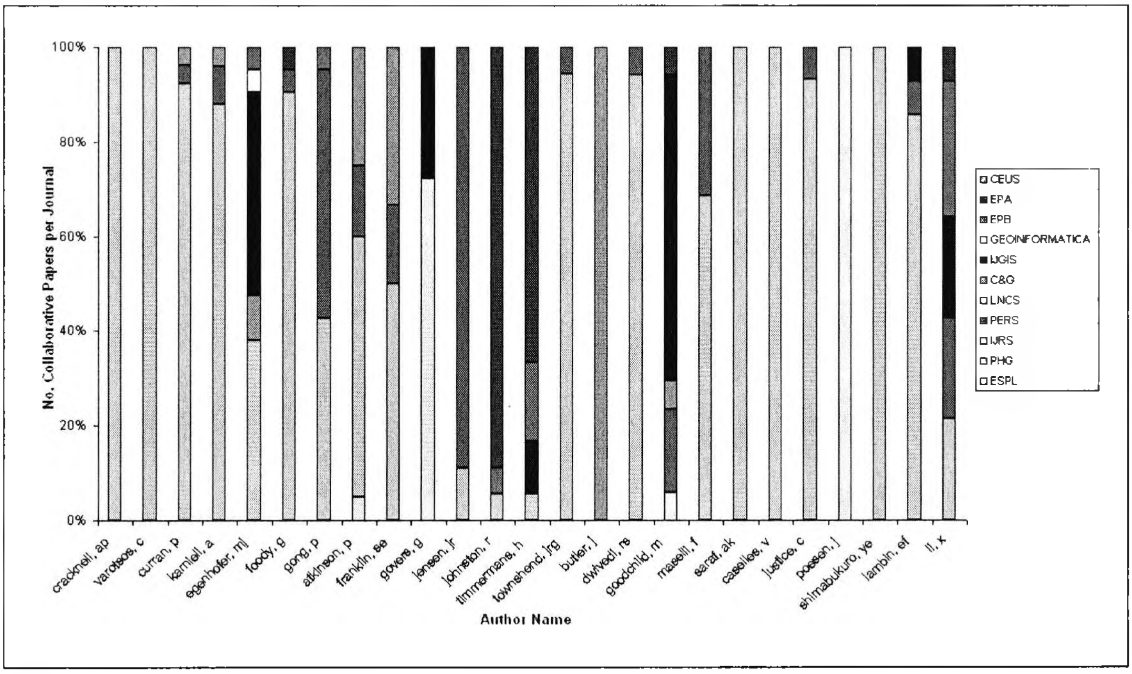


Figure 5-27 Journal distribution of top-network author participation through collaborative works. The figure shows 25 authors as Caselles, Justice, Poesen, Shimabukuro and Lambin the top less participative authors (on the right hand side) have all 15 collaborative papers.

Therefore, one could label them as specialized remote sensing scholars within the context of this co-authorship network. In the same way, C. Pattie specializes in environmental issues and J. Butler in GIScience core topics. However, there are other groups of authors that observe more multidisciplinary research backgrounds as they participate with works on more than one of the research topic groups. For example, through their contributions, one can say that some authors research on environmental and work on any geoscience or on core GIScience issues. Here, it is important to note the place of journals such as LNCS, IJGIS, Geoinformatica and C&G (shades of blue) as general forums for authors covering not only work on very specify areas, but also publishing findings on core topics of the discipline.

From this very participative and multidisciplinary group in Figure 5-27, the presence of M. Goodchild stands out. His collaborative participation pattern shows an author whose work is not concentrated but is spread around a variety of topics concerning the discipline. Additional, authors such as P. Fisher, D. Martin and AS. Fotheringham (see Figure 5-26) with a less collaborative participation, have a similar multidisciplinary pattern to M. Goodchild. As chapter 4 identified some of these authors are key players in building the co-authorship network, but also they are well-known scholar within the research community.

Until now, this section analysed the degree of an author's contribution in terms of research topics. The next section focuses on analyzing the multidisciplinarity of the scholar-scholar connections as a result of joint contributions.

5.3.2 Multidisciplinarity in GIScience Scientific Collaborations

As the topology of a co-authorship network is built upon collaborative papers, it is important to study not only to which degree authorship is spread over all covered research topics (journal topic), but also the nature of the author's co-authorship links. In doing so, this section concentrates on exploring the multidisciplinary nature of the co-authorship links from the top-most active scholars (studied in the previous section) and from the scholars who have works with the largest number of collaborators.

This study assumes that the general topic of the journal, in which a collaborative work is published (see broad research topical groups in section 5.1.2) can be taken as representative of the research interests of all participating co-authors. Hence, for each collaborative paper, an author has (*total number of co-authors – 1*) co-authorship links in the given journal subject. The procedure is repeated for each one of the joint works of the top-participative authors listed in Figure 5-26.

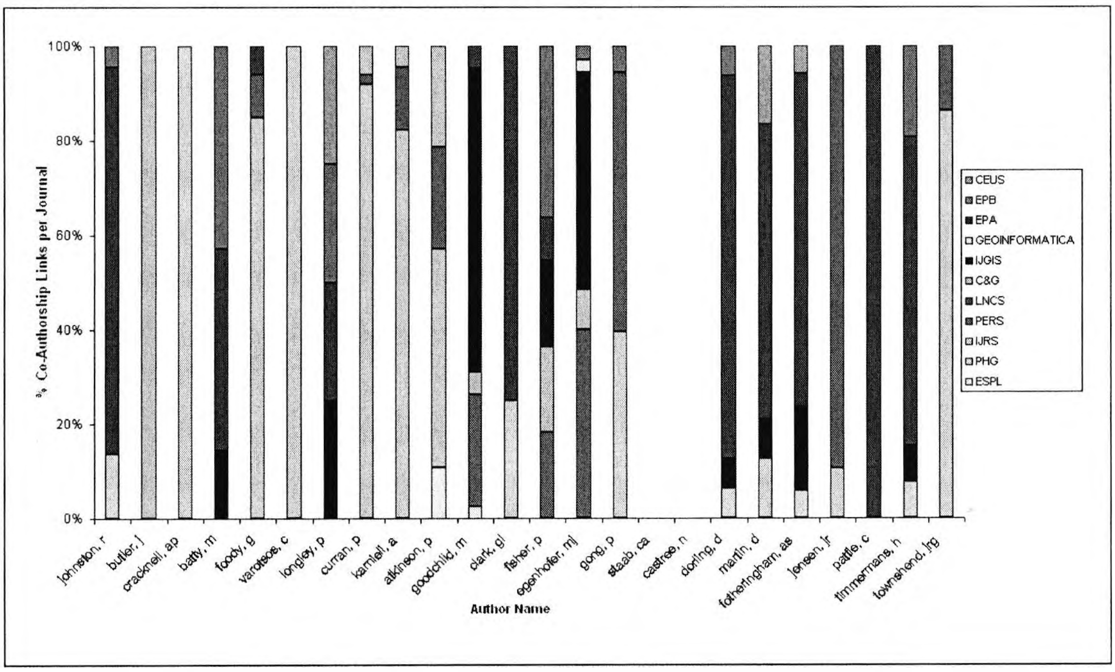


Figure 5-28 Distribution of the topical backgrounds of co-authorship links from the joint works of the top participative scholars (shown in Figure 5-26).- Observe that CA. Staab and N. Castree do not have co-authorship links as their participation is only through sole works.

Figure 5-28 shows that authors such as AP. Cracknell, JR. Jensen or C. Varotsos who publish only on remote sensing, have only remote sensing collaborative links. On the contrary, authors who diversify their publication sources such as M. Goodchild, MJ.

Egenhofer or P. Fisher exhibit a multidisciplinary set of co-authorship links. Besides, the figure shows a strong presence of links from remote sensing and environmental issues within the top-author list, a consequence of a high average of authors per paper (see Table 5.33). Moreover, as each shaded bar height in Figure 5-28 can be interpreted as a relative measure of the collaborative papers in term of no. co-authors, one can see the difference in size of remote sensing (shades of pink), environmental (shades of green) or GIScience (shades of blue) research teams in which top-scholars participate.

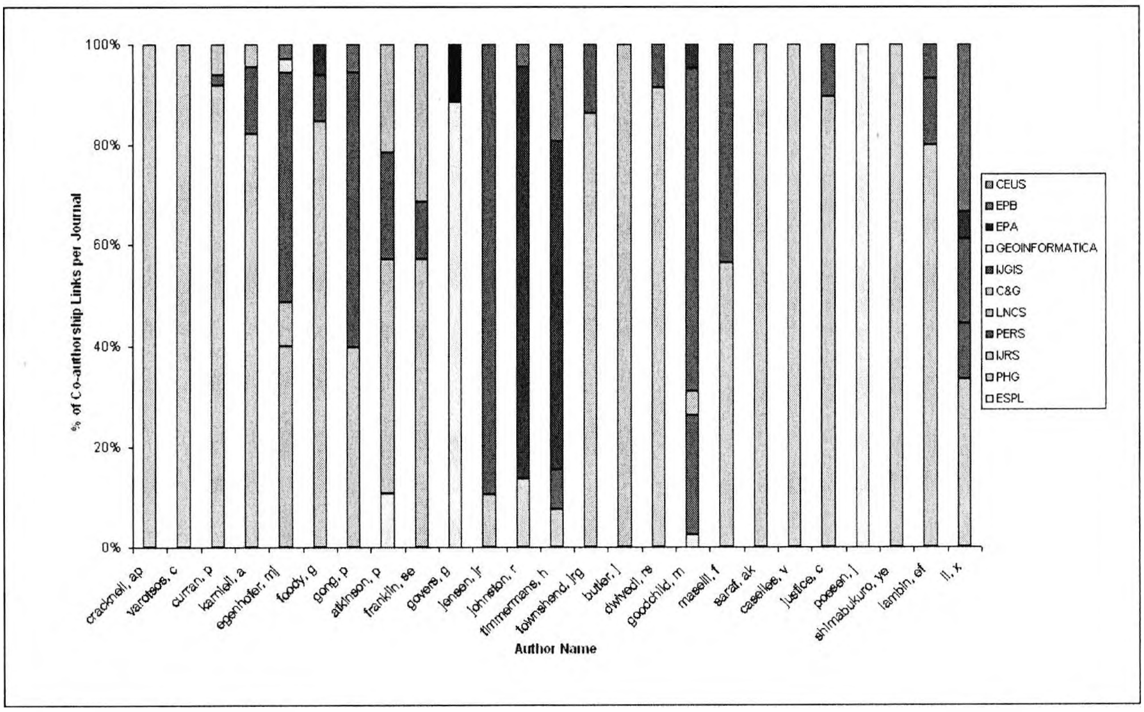


Figure 5-29 Distribution of the topical backgrounds of co-authorship links from authors with the highest number of joint works shown in Figure 5-27. Authors with one shade have all his/her collaborators from one journal only.

However, the list of top collaborative authors changes if only joint works are taking into account. The distribution of scientific links in Figure 5-29 from the new top-authors (the same authors listed in Figure 5-27) reveals a much stronger dominance of remote sensing. The figure shows that 10 out of 25 top collaborative authors have joint publications mainly in remote sensing journals (see Figure 5-27). Contrasting, M. Goodchild and MJ. Egenhofer have a much diversified set of authorship links from many of the research topics covered. Hence, they can be taken as representative of the multidisciplinary nature of scientific collaboration on GIScience core research.

Focusing on authors with the highest number of co-authors, Figure 5-30 reveals that remote sensing is also predominant as the background from authors who have written their works with the largest number of collaborators.

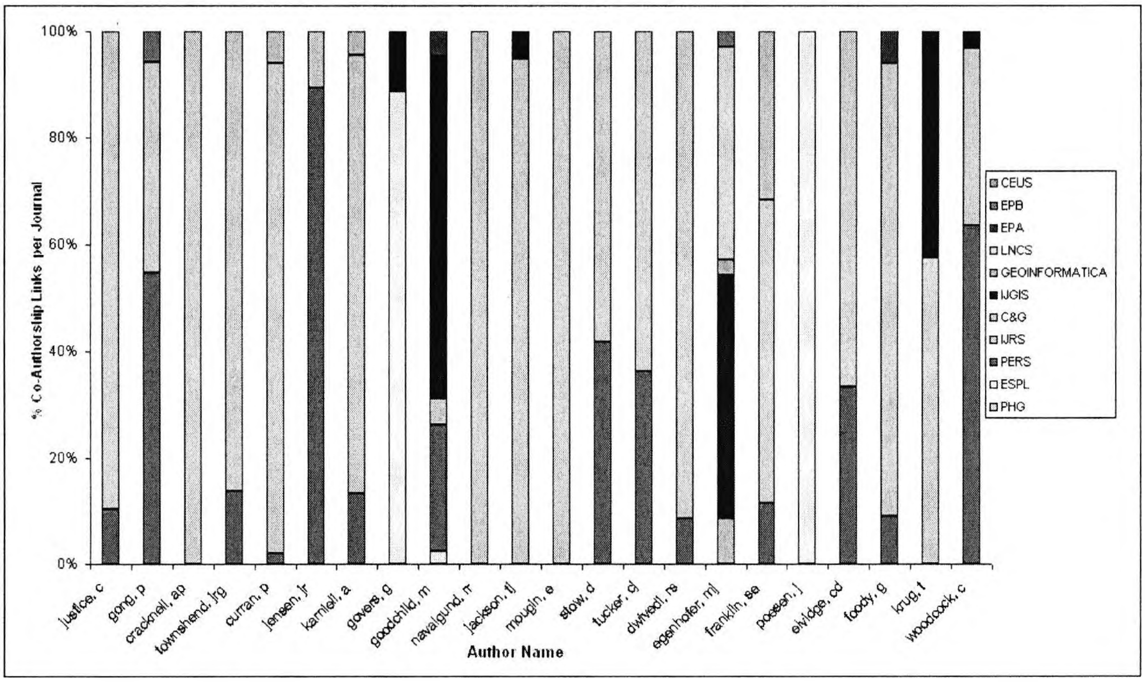


Figure 5-30 Distribution of the topical backgrounds of co-authorship links from the top 20-authors with the highest number of collaborators. The figure listed 22 authors as T. Krug, C. Woodcock, CD. Elvidge and G. Foody have all 33 co-authorship links.

Using top-authors regarding participation and collaboration, the previous two sections show that the network multidisciplinarity depends on the collaborative nature of the works (sole or joint) and also, on the multidisciplinary nature of the links between participant co-authors. The results show that some authors despite their high levels of participation do not make much impact on the network topology because the majority of their works are individual ones. Besides, the majority of joint works of top authors (regarding number of joint and collaborators) show that their works are concentrated on few research issues. Thus, one can conclude that with the exception of very few, top authors exhibit a behaviour that is not oriented towards a multidisciplinary co-authorship. Moreover, that the dominance of remote sensing as a research topic (IJRS and PERS), where authors seem to publish more works and with more co-authors than the average network, is evident, resulting in a network topology heavily influenced by them and their collaborations. Some authors such as M. Goodchild, D. Martin and MJ. Egenhofer exhibit the opposite multidisciplinary co-authorship behaviour that better describes the multidisciplinary nature of research on GIScience (see section 5.1).

However, one cannot describe the multidisciplinarity of a co-authorship network based only on the co-authorship and collaboration patterns from the top-authors, because they are representative of less than 1% of network participants. Moreover, participating with a high number of joint works does not mean a highly collaborative author (see R.

Johnston). In the same way, having a large number of collaborators does not indicate a very participative author. For example, an author may have 10 joint works co-authored with one collaborator each time. At the same time, an author with a high number of co-authorship links cannot be labelled as very participative, as they may come from one or two joint papers (see 68 author-paper in section 4.1.2- chapter 4). Hence, the next section explores how to quantify the multidisciplinary nature of authors within the network and to analyze the concentration of authorship. This will help determining the multidisciplinary nature exhibited by the network.

5.3.3 Multidisciplinary Characterization of GIScience Core Network

Characterizing multidisciplinary in a discipline such as GIScience where many other disciplines, fields and areas may overlap, is not an easy task. The task becomes more difficult as it is not a normal procedure for bibliographic databases such as ISI-WoK individually to index each published work according to its specific topic (Leydesdorff, 2006). Instead, they categorize all papers under the topic of the journal, in which the work is published. For example, regarding journals where applied GIScience research may be published such as CEUS, ESPL, PERS or PHG, one cannot say that all work published there is related by some means to GIScience. Thus, as Bensman (2001) states, an ambiguous categorization of the journal set for a discipline, specially one as highly multidisciplinary as GIScience, in terms of subject matters seems impossible because of the fuzziness of the subsets. To overcome this problem, Leydesdorff (2006) proposes a betweenness centrality as indicator of interdisciplinary of scientific journals. However, this measure is based on how many times journals cite one another, hence, it is not suitable as this study focuses on co-authorship. A simple inspection of authorship patterns from highly influential authors gives a very broad and general idea on how co-authorship is spread. However, as the network is comprised of around 14000 authors, it is important to find a measure that quantifies authorship homogeneity in the 11 selected journals.

In traditional thermodynamics, entropy is a measure of the amount of energy in a closed system (the second law of thermodynamics) that is no longer available to effect changes in that system (Carrier, 2005). Thus, as Carrier (2005) explains, entropy changes are used to measure how much energy is spread out in a particular process, or how widely spread out it becomes. In 1948, Claude Shannon connected information theory and physics by developing a new perspective on entropy related to the second law of thermodynamics. Shannon (1948) used entropy to quantify the information content of a source part of a communication channel. The Shannon's information entropy of a discrete variable *x* is

$$E(I(x)) = \sum_{i=1}^{n} p(x_i)\log_2\left[\frac{1}{p(x_i)}\right] = -\sum_{i=1}^{n} p(x_i)\log_2 p(x_i)$$

Equation 5-6 Shannon's Information Entropy

Where *I(x)* is the information content or self information of x, and $p(x_i)$ is the probability mass function of *x*.

5.3.3.1 Entropy (E) and Scaled Entropy (SE)

For a categorical classification problem Wood (1996) used the scaled entropy (SE), a derived measure of Shannon's information entropy (defined in Equation 5-6), to quantify the degree of variability of a location's classification regarding the scale. The variation is measured by the Entropy (E), in this case defined as:

$$E(x_{i)}) = -\sum_{i=1}^{n} n*(p_i * \ln[p_i])$$

Equation 5-7 Entropy

Where *n* is the number of categories in which the variable is spread out, and *p* is the proportion of values in each category. To scale *E* values between 0 and 1, Equation 5-8 is divided by the maximum entropy possible according to the existing categories. The maximum entropy *(MaxE)* is defined as:

$$\text{MaxE} = -\text{n}*\left(\frac{1}{\text{n}}\right)*\left(\ln\left[\frac{1}{n}\right]\right) = -\ln\left[\frac{1}{n}\right]$$

Equation 5-8 Maximum Entropy

Where *n* is the number of categories. Therefore, the scaled entropy (SE) is defined as:

$$SE = \left(\frac{E(x_i)}{MaxE}\right)$$

Equation 5-9 Scaled Entropy

Using SE, Wood (1996) was able to distinguish locations that are consistently classified as the same feature (SE = 0) from locations that have a high degree of scale dependency in their classification (SE = 1). Similarly, but within a more similar research context to this study, Börner *et al* (2005) proposed SE to measure whether the impact (a weighted measure based on the number of scientific publications and their respective citations) of an author is spread evenly over all her/his co-authors. A given author exhibiting SE = 1 shows that all his/her collaborators hold a similar impact measure. If SE = 0 then a single scientific collaboration accumulates more weight than the others, and so is more likely to make a higher impact.

Consequently, this study proposes to use SE (see Equation 5-9) as a measure of the dispersion of authorship across the 11 journals that comprise the co-authorship network. If the majority of authors exhibits SE = 1, then the network comprises a group of scholars with a very multidisciplinary research interests. Otherwise, if the majority of authors observe SE = 0 or close to 0, the network is portraying a group of authors working in more specific research areas. In other words, one journal is the unique publishing forum for all papers from the majority of authors. One can conclude that authors showing SE closer to 1.0 are those doing multidisciplinary research, as their publications are more evenly spread across a variety of GIScience or related topics represented by the included journal set.

5.3.3.2 Co-Authorship Concentration

The SE measures offer a way to corroborate some of the findings presented in chapter 4 regarding the network topology. In order to interpret SE results in the context of the GIScience multidisciplinary research represented by the co-authorship network, one needs to revisit first some of the factors that influence the network per se. One of these factors is the percentage of authors participating in the network with one paper only. In total, almost 90% of the authors exhibit authorship concentration (SE=0). However, from this total around 79% of the authors have the observed concentration because they have only one work within the network, leaving only 21% of the authors that exhibit authentic authorship' concentration. Therefore, to avoid misleading conclusions, when SE =0, the concentration of authorship is analyzed into two separate groups. One that is represented by authors who participate with only one paper, and a second group of authors with more than one paper.

The former group represents authors of whom one cannot say that they may have or not multidisciplinary research interests as their network participation is very low. Figure 5-31 shows that 47% of these authors publish once on remote sensing topics (pink shades), 18% on environmental (green shades), 13% on geosciences (yellow shades) and finally, 22% from core GIScience and related computer issues (blue shades).

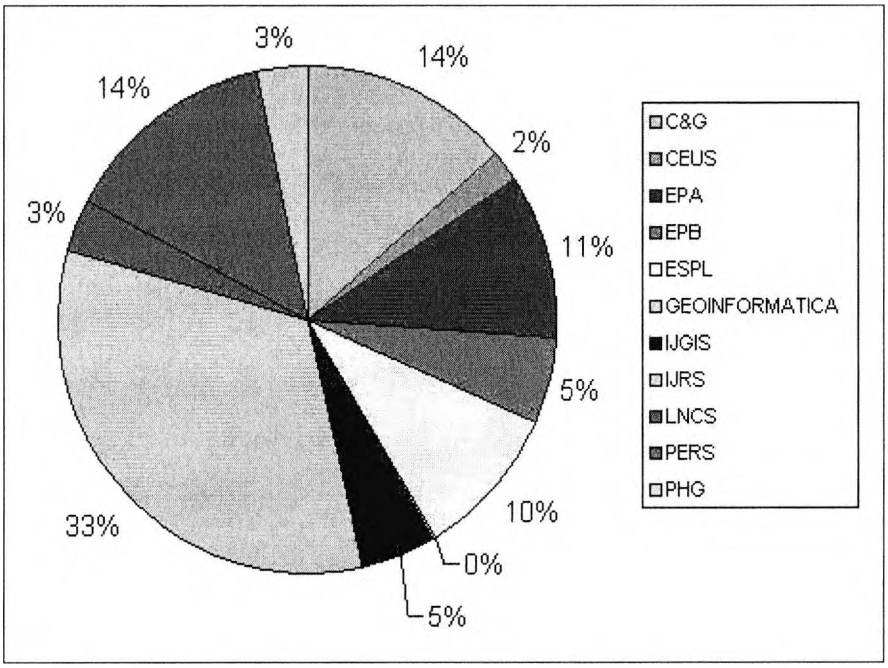


Figure 5-31 Distribution of the authorship concentration (SE=0) when authors participation = 1. Despite of having SE=0, authors do not exhibit authorship concentration as they have published one paper only.

The latter group comprises authors with at least two papers in the same journal. Thus, one can label them as having research interests concentrated on that specific journal topic.

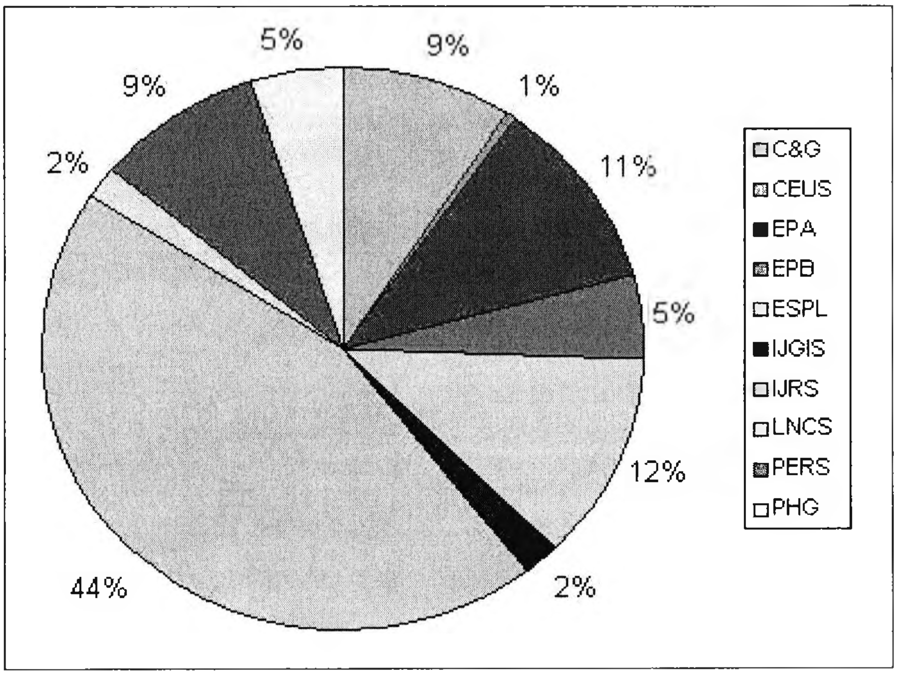


Figure 5-32 Journal distribution for the authors that observed authorship concentration (SE =0 and authors' participation > 1).

Within this group, Figure 5-32 reveals that 53% are authors working only on remote sensing, 17% correspond to environmental and geosciences authors. Finally, 13% of authors in this group are authors with research interests concerned with GIScience core topics. According to Table 5.35, J. Butler, stands out as a highly participative author (53 papers) with core GIScience topics as his main research interest.

Author Name	Total Network Participation (sole and joint works)	Research Interest concentrated on	Author Name	Total Network Participation (sole and joint works)	Research Interest concentrated on
anderson, nl	11	C&G	rees, wg	19	IJRS
butler, j	53	C&G	ricotta, c	11	IJRS
loudon, tv	14	C&G	saraf, ak	19	IJRS
wallace, r	13	EPA	shimabukuro, ye	15	IJRS
abrahams, ad	11	ESPL	singh, rp	11	IJRS
poesen, j	15	ESPL	vandermeer, f	11	IJRS
askne, j	12	IJRS	varotsos, c	31	IJRS
cartalis, c	12	IJRS	mugnier, c	11	PERS
caselles, v	15	IJRS	mugnier, cj	18	PERS
gupta, rp	12	IJRS	staab, ca	23	PERS
kondratyev, ky	11	IJRS	dodds, k	11	PHG
oppenheimer, c	11	IJRS	watts, m	11	PHG

Table 5.35 The most participative authors that observe authorship concentration SE = 0, then focused research interests.

In the same way, all participation from C. Varotsos and Ca. Staab come from work on remote sensing. Therefore, one can assume that authors listed in Table 5.35 have as their research focus the topic of the journal where they have concentrated their publications. It is important to observe that within both groups, the percentage of authors with papers in Geoinformatica and SE = 0 is almost zero respectively. It means that there is almost non co-authorship concentration of authors in this journal. From the total of 69 authors (see author totals in Table 5.33) with works in Geoinformatica, 31 authors participate with only one paper, and not one author has published twice in this journal (as there are not authors with SE = 0 and more than one paper from this journal). This implies that the remaining 38 authors have works published in at least one of the other participant journals.

5.3.3.3 Co-Authorship Distribution

The 11% percent of participating authors observe a more distributed set of works across all journals (SE > 0). However according to Figure 5-33, SE values are not higher than 0.50 for authors whose publication forums comprise less than 3 journals. Moreover, Figure 5-33 reveals that some authors despite publishing in 4 journals have low SE values (SE < 0.3), indicating a less evenly authorship distribution. On the contrary, authors with the highest SE values (SE > 0.6) are those whose works are more evenly distributed among the highest number of journals, in this specific case in more than 5 journals.

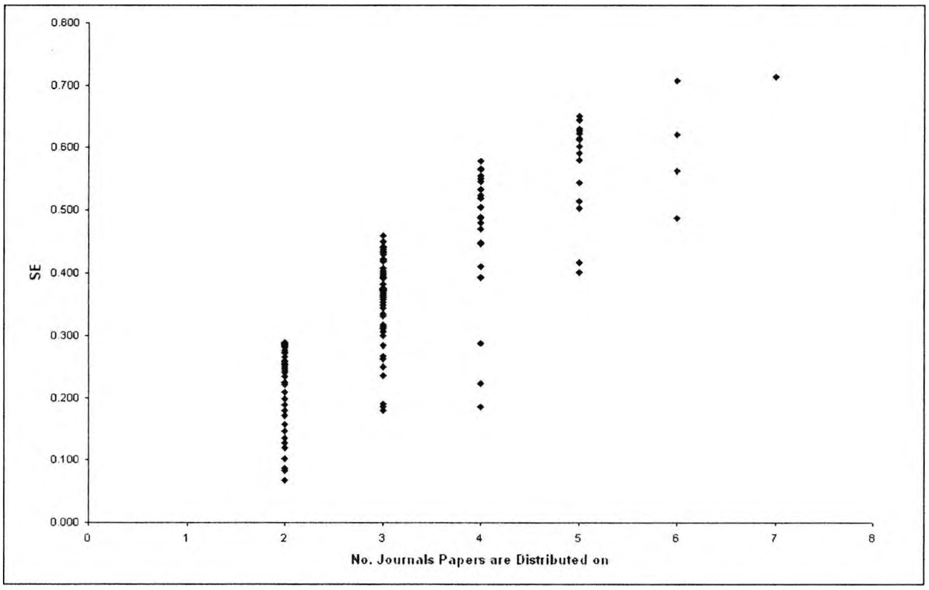


Figure 5-33 SE Spectrum versus Journal Distribution for authors with SE > 0 and Network Participation > 0.

Therefore, two different patterns can be distinguished among the authors with SE > 0. One refers to authors who despite having distributed their publications do so among a very few number of journals. The second group corresponds to authors who have more evenly distributed their publications in the greater number of journals (SE > 0.60).

Regarding the former group, 96% of the authors show some degree of spread (SE< 0.60). Within them, 86% have their work distributed among two journals, and around 15% among 3 or more journals (see Table 5.36).

No. Journals	No. Authors
6	2
5	9
4	41
3	227
2[8]	1264

Table 5.36 Journal distribution for authors with 0.0 < SE <0.60 (network participation > 2 papers)

If one excludes authors with two papers (as the two are equally distributed among 2 journals), Figure 5-34 reveals that 37% authors with some degree of spread published mainly in remote sensing journals. In the same way, around 30% have published in environmental journals and around 17% in both, GIScience and geosciences journals.

[8] Here, 2 journals do not mean that the authors have only two works. They can have many papers that have been published between the two journals.

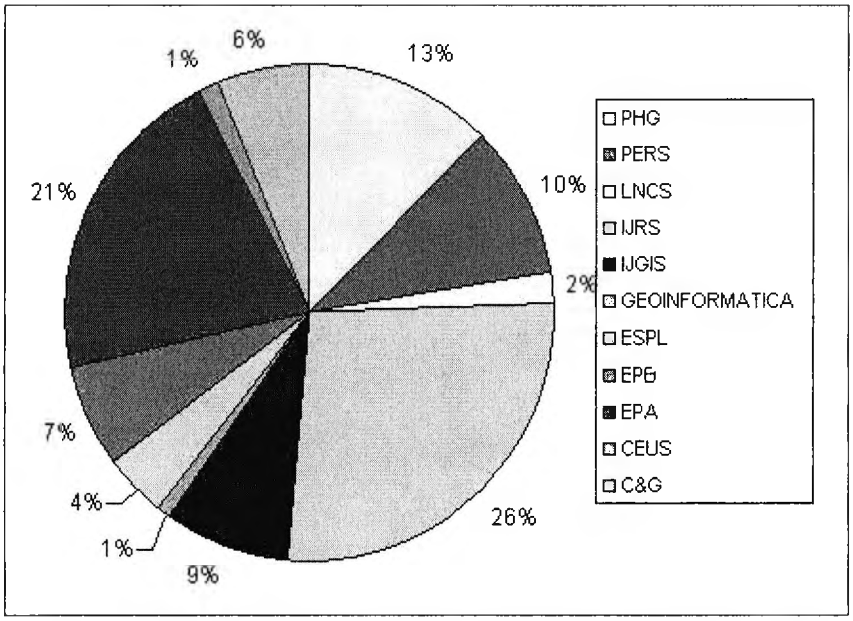


Figure 5-34 Distribution of journals with the highest proportion of works from authors exhibiting some degree of distribution (SE<0.60 and participating > 2 papers).

However, because multidisciplinarity has been associated with GIScience, the central point is to identify authors that have distributed their work on the highest number of journals. Table 5.37 displays authors, that despite having SE < 0.60, have publications in at least 5 of the 11 participating journals. Then, they can be labeled as portraying very multidisciplinary research interests.

Author Name	Network Participation (sole and joint works)	SE	Journal with the highest proportion of works	Total No. Journals paper are distributed in
egenhofer, mj	23	0.563	IJGIS	6
goodchild, m	25	0.487	IJGIS	6
atkinson, p	25	0.514	IJRS	5
foody, g	32	0.401	IJRS	5
fotheringham, as	20	0.503	EPA	5
jankowski, p	10	0.592	IJGIS	5
lo, cp	10	0.592	PERS	5
longley, p	30	0.417	EPB	5
martin, d	21	0.580	EPA	5
openshaw, s	18	0.401	EPA	5
wise, s	9	0.543	IJGIS	5

Table 5.37 List of authors with the highest SE values in the first group (0.0<SE<0.60), their network participation and their most frequent publication forum.

Table 5.37 also shows the journal in which each author has published the highest number of papers, it is important to note a less strong presence of remote sensing authors in this list. As the results from previous sections have shown, one can conclude that this is due to the fact that despite being highly participative and collaborative authors, their works are mainly distributed on the two remote sensing journals.

Finally, the last group comprises authors whose work is not only more evenly distributed, but is in a greater number of categories (SE > 0.60). Table 5.38 displays the full list of this truly multidisciplinary set of authors.

Author	Total No. Papers	SE	Journal with the highest proportion of papers	No journals that papers are distributed among
fisher, p	24	0.713	EPB	7
gopal, s	10	0.707	IJRS	6
arentze, t	6	0.651	IJGIS	5
smith, a	6	0.651	EPA	5
treitz, p	6	0.651	PERS	5
armstrong, mp	11	0.645	C&G, CEUS	5
li, x	16	0.630	EPB	5
lin, h	10	0.628	IJGIS	5
wang, f	8	0.623	EPA	5
lee, j	8	0.623	PERS	5
frank, au	17	0.620	IJGIS, LNCS	6
clarke, kc	7	0.615	IJGIS	5
campbell, h	10	0.613	EPB	5
masser, i	10	0.613	IJGIS	5
gao, j	14	0.613	IJRS	5
yeh, ago	12	0.603	EPB, PERS	5

Table 5.38 Authors with the highest SE, their network particiaption, journal with the highest proportion of work, and the total number of journals they have published in.

According to the figures in Table 5.38, in most cases the higher the SE of an author, the higher the number of journals among which his/her work is distributed. Moreover, authors with SE > 0.60 show their work distributed across at least 5 journals. It is important to note that a higher SE does not imply a highly collaborative author. For example, P. Fisher's collaborative work represents only 20% of his network participation. In contrast, 90% of S. Gopal's participation is collaborative.

The distribution of SE versus the collaborative network participation, in Figure 5-35, reveals a group of authors with different degrees of SE who tend to publish individual papers in more than 2 journals (y-axe). The distribution also reveals that authors who distribute their works in the higher number of journals (SE > 0.60), tend to participative in teams smaller than 10 authors.

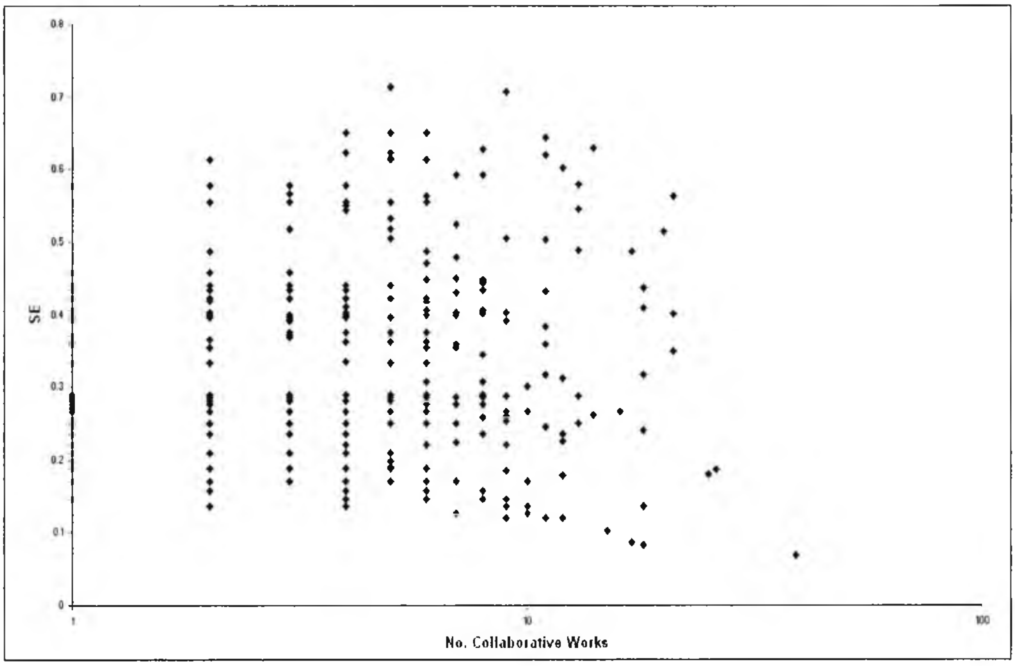


Figure 5-35 SE spectrum versus the Distribution of Collaborative Work (X-Log Plot)

This argument is corroborated by plotting the spectrum of SE against the authors' network participation. The distribution in Figure 5-36 shows that one can find either authors participating with less or more than 10 papers that hold high SE values, or authors with more than 15 papers but computing a very low SE.

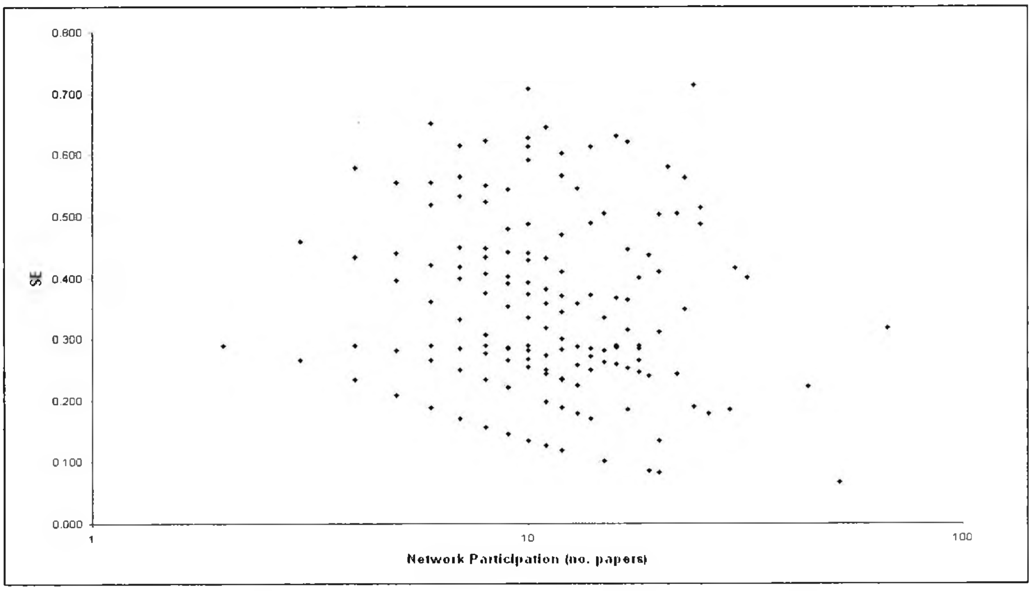


Figure 5-36 SE Spectrum versus network participation (X-Log Plot)

It reveals that the former authors' work is more evenly distributed in a greater number of journals or, better, that authors' research interests focus on the majority of topics covered by the network (see P. Fisher or S. Gopal). On the contrary, the latter shows authors whose works are mainly focused on few topics, but also contribute in less proportion to other topics. In this sense, figures in Table 5.38 show that authors have a journal of preference where a highest proportion of their works are published. Despite multidisciplinary research interests, one can say that this particular journal may reveal the research focal point of the author. There is not a unique focal point among the small group of authors with SE>0.60. Instead, they seem to represent all the topics covered by the network.

This chapter reveals a co-authorship collaboration network strongly influenced by remote sensing publication outlets. One can say that the large number of papers published in remote sensing journals may account for it. However, it is also a consequence of the highly collaborative and participative nature of the authors publishing on this topic. By examining the top participative author list is revealed that they are mainly working in remote sensing. It shows also that these authors do not exhibit a homogenous participation, but rather a high concentration of work on the two remote sensing journals. Moreover, the large part of their collaboration comes from the work published on those topics. Though, remote sensing authors dominate the network topology, the study found other authors that exhibit a more multidisciplinary nature closer related to GIScience research. The majority of these authors have at least 3 research focuses and around 10 collaborators.

6 The Geography of GIScience Scientific Collaboration Networks

Previous chapters focused on examining the topology, centrality measures and the multidisciplinary nature of the GIScience collaboration network from a statistical and mathematical viewpoint. This chapter analyzes not the statistical but the geographical distribution of the collaborations at local (individual) and global (national) levels. In doing so, one will be able to geographically characterize the network using the spatial distribution of co-authorship linkages and to determine whether or not there is any geographical trend. The network's spatial characterization task involves a threefold process: firstly, to explain the geo-referencing process including the limitations imposed by the bibliographic database. After understanding the problem ahead, the second step involves geo-referencing the non-spatial co-authorship network through the individual spatial locations of its co-authors. The georeferencing process entailed the development of a computerized tool that helps to extract geographical locations and to create the respective geo-referenced scientific collaboration pairs. The final step examines whether or not one can identify any facilitator such as language or geographical proximity at the time of collaboration. All three steps are elaborated within the following sections.

6.1 Geographical Component of the GIScience Collaboration Network

Statistical properties of the GIScience collaboration network (see section 4.1 Chapter 4) revealed similar results to studies on co-authorship networks from various disciplines (Ding *et at*, 1999; Barabasi *et at*, 2000; Newman, 2001a; Horn *et at*, 2004; Moody, 2004; Newman, 2004b; Newman, 2004c; Elmacioglu and Lee, 2005; Liu *et at*, 2005), in which the distribution of collaboration links are highly skewed, with a small number of scholars having a large number of co-authors, while most have very few. However, none of these studies has taken into account the spatial component of the co-authorship network that can be generated by the geographical locations of a paper's co-authors. This section elaborates on how the spatial component of the GIScience authorship network was added, limitations encountered regarding the bibliographic data used and an insight into the creation of the actual maps.

6.1.1 Geo-referencing and Co-authorship Networks

Within the GIScience realm geo-referencing, according to a very condensed definition, refers to a process of referencing in space any point using a predefined coordinate system such as a national grid or latitude/longitude (Longley *et al*, 2001). Thus, one widely spread method of geo-referencing a co-authorship network (Wagner, and Leydesdorff, 2003; Glänzel and Schubert, 2004; Arciniegas, and Wood, 2006; Calero *et al*, 2006; Carvalho and Batty, 2006; Murray *et al*, 2006; Leydesdorff, 2006a) consists of assigning geographical locations to all participant co-authors.

Among these works, Batty (2003) and Batty (2003a) incorporated the geo-spatial component and used the most highly cited authors in 2002 from ISI-WoK database, revealed that different levels of spatial aggregation of authors, despite their scales, tend to show that fewer institutions, research centres or countries attracted the highest number of citations. For example, at institutional level such as research centres or university departments, citations are concentrated on few scholars affiliated to few university departments or institutions.

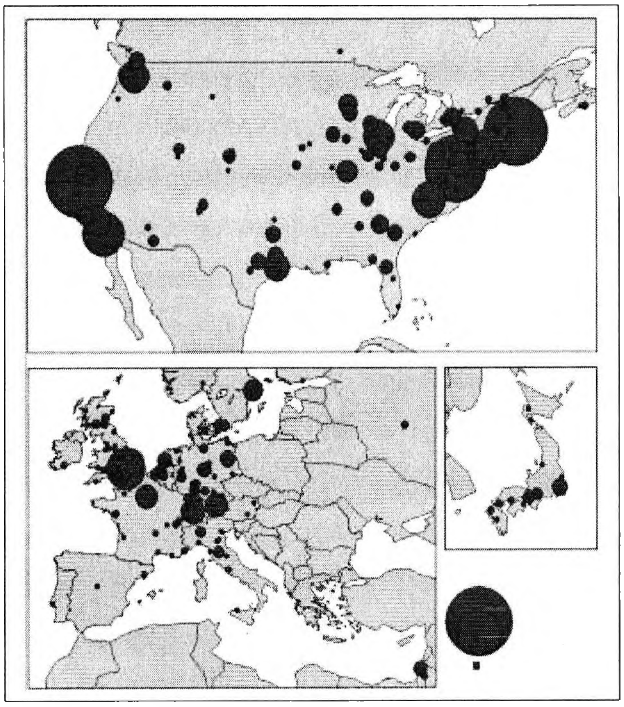


Figure 6-37 The Geographical distribution of highly cited authors during December 2002. This study excludes mathematics, social science and humanities (extracted from Batty, 2003a, pp. 764).

From Batty (2003a)'s results (see Figure 6-37), it is evident the high impact within their respective communities of works from institutions around Boston, New York, California and central London area. As Batty (2003) states, these locations bear out the perceptions of where the world's top institutions are most heavily concentrated. However, this study substantially differs from the present one, as it focuses on citation instead of co-authorship measures.

Shifting the focus to scientific collaborations, Calero *et al* (2006) used authors' affiliation data from nanotechnology publications indexed by ISI-WoK to identify centres of research excellent in that field. Wagner and Leydesdorff (2003) studying the global science collaboration network, revealed that the network core, comprised of industrialized nations, has expanded over the years encompassing new nations. In spite of both studies having the research centres' addresses and countries' names as geo-referencing points, their results are presented using the network graph. As a consequence, for the less common countries of which many are unaware of their locations, the importance of the geographical proximity is not easily appreciated, but this would be highlighted if a map were used.

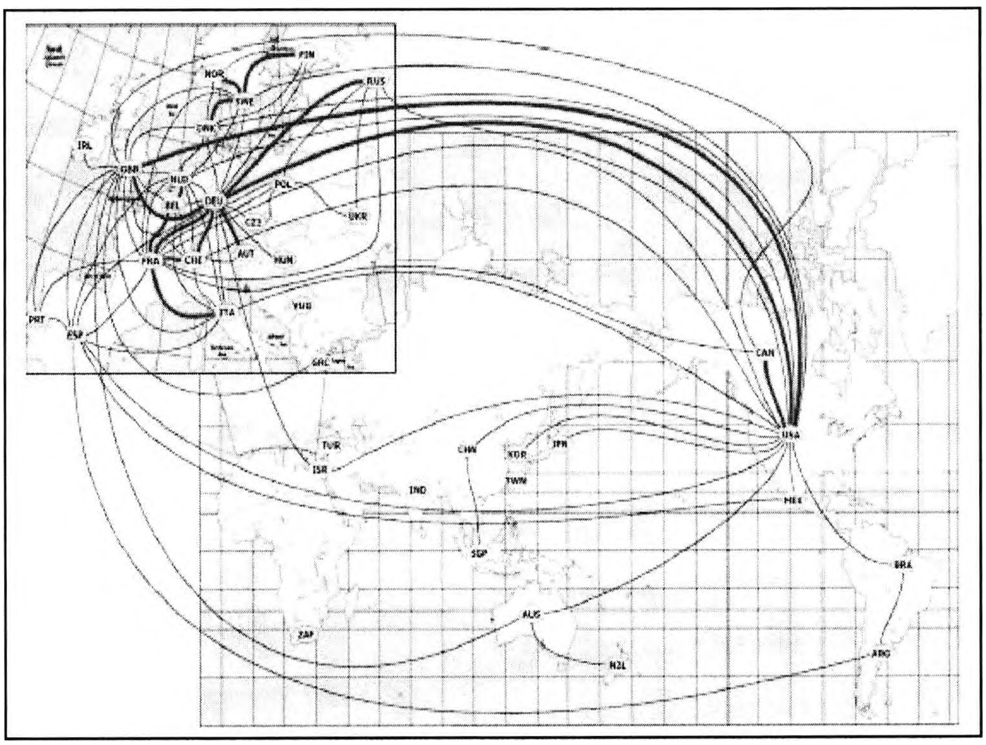


Figure 6-38 Co-Authorship map for the most active countries in all fields during 2000, based on Salton's Measure (SM). Dotted line corresponds to SM>= 1.0%, solid line SM>= 2.5%, thick line SM >= 0.5%. (Glänzel and Schubert, 2004, pp. 268).

Examining global collaboration in science, Glänzel and Schubert (2004) used the Salton Measure (SM) to compute the strength of collaboration between a pair of countries during the year 2000.

The SM is defined as

$$SM = \frac{n_{ij}}{\sqrt{n_i * n_j}}$$

Equation 6-10 Salton Measure

Where n_{ij} is the total output of joint publications between *countries i* and *j*, n_i is the total number of publications for *country i* and n_j is the total number of publications for *country j*. Within a global perspective, Glänzel and Schubert (2004)'s results for the year 2002 (see Figure 6-38) revealed that besides UK, USA and France, Germany is also one of the worlds most important research nodes. In addition, the map in Figure 6-38 reports a highly connected cluster among Scandinavian countries (Denmark, Sweden, Norway and Finland) scientific relationships had became stronger, if compared to Glänzel and Schubert (2004)'s results for the years 1980 and 1990.

Focusing specifically on GIScience, Arciniegas and Wood (2006) extracted co-authorship data indexed by WoK from 11 GIScience journals, and geo-referenced the set using authors' affiliation information.

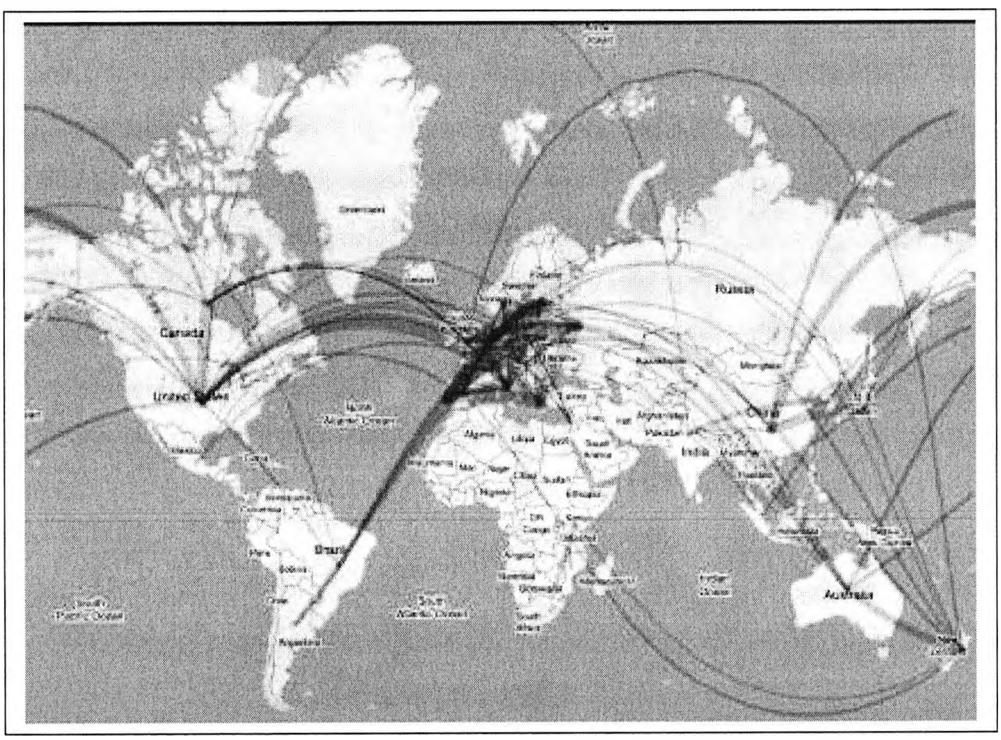


Figure 6-39 International collaboration links in GIScience from papers published in 11 relevant journals during 2002. The map shows all country collaborations where n_i and n_j are both greater than 10 links and e_{ij} (exclusivity index between a country pair) is at least 2% (Arciniegas and Wood, 2006, pp. 553).

They calculated a collaboration exclusivity index to quantify the strength of the linkage between a pair of countries. The results of aggregating the bibliographic data at country level for 2002, displayed in Figure 6-39, pinpointed strong collaboration links between countries with English as a native language. Moreover, a strong relationship can be seen between European countries. Brazil and Argentina can be considered as emerging countries within Latin-American. Similarly to Glänzel and Schubert (2004) and during the same year, Arciniegas and Wood (2006) identified a strong relation among Scandinavian countries. Also a very strong partnership was identified between Russian, Finish and Greek scholars, as a result of six joint publications during this year by authors affiliated in these countries.

However, a one year time window used by Arciniegas and Wood (2006) may lead to erroneous conclusions as the linkages can be the result of what Glänzel and Schubert (2004) identified as temporary or occasional links. Thus, the time window needs to be expanded in order to determine if these are regular patterns instead of sporadic connections. Therefore, this study concentrates on further analysis of the spatial distribution of GIScience scientific linkages during the same 11 year time window, as covered by the statistical analysis. The following section details the problems encountered while geo-referencing the data, and lists all assumptions that were required.

6.1.2 ISI-WoK Limitations and the Geo-referencing Process

All abovementioned analyses of geographical distributions of scientific collaborations have used the ISI-WoK as bibliographic data sources. As fully explained in chapter 3, ISI-WoK offers the possibility of saving any query results into text files. Apart from the most traditional information regarding bibliographic documents such as title, authors, and keywords (see Table 6.39), each bibliographic record saved from ISI-WoK contains useful information referring to the co-authors' addresses under the C1 and RP labels in the example represented in Table 6.39. Therefore, it is C1 and RP information that have allowed bibliometric studies to geo-reference authors of scientific documents (see grey rows in Table 6.39).

At the time of their publication, Bordons and Gomez (2000) advocated that ISI-WoK was one of the few comprehensive information databases that stores author addresses alongside the publications' details. Though the ISI-WoK exhaustiveness is not called into question (Glänzel and Schubert, 2004; Batty, 2003; Calero *et al* 2006), personal experience acquired while working with the bibliography data set drawn from ISI-WoK, suggests that some aspects concerning the way authors' addressees are indexed hinders the author geo-referencing process (this is referred back in July 2005, when the bibliographic data for this study was downloaded from ISI-WoK database).

PT	Journal
AU	Cova, TJ; Goodchild, MF
TI	Extending geographical representation to include fields of spatial objects
SO	INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE
DT	Article
ID	INFORMATION-SYSTEM; CELLULAR-AUTOMATA; GIS; MODELS
C1	Univ Utah, Dept Geog, Salt Lake City, UT 84112 USA; Univ Calif Santa Barbara, Dept Geog, Santa Barbara, CA 93106 USA
RP	Cova, TJ, Univ Utah, Dept Geog, 260 S Cent Campus Dr,Rm 270, Salt Lake City, UT 84112 USA
ER	End of Record

Table 6.39 An example of a record retrieved from the ISI-WoK)

The argument is better understood through an example. The bibliographic record in Table 6.39 represents a journal paper written by two authors published in IJGIS in 2002. It shows two addresses that one assumes correspond to each one of the paper's co-authors. In order to georeference the co-authors, one would assign each address to each author. Though here, the process of pairing author addresses seems straightforward, according to ISI-WoK documentation in 2005 one cannot be sure that the first address corresponds to the first author, or the second to the second author. Moreover, two others cases were observed, one were a record comprises more authors than addresses and two, when there are more addresses than authors. Thus, the lack of agreement between the number of authors and the addresses indexed in each paper makes it difficult to georeference the network at the author level. Consequently, geo-referencing each of the participant authors could not be carried out, and therefore, an alternative method was devised.

The proposed approach to georeference the network entails aggregating data by country, sacrificing authors' identities. Thus, the countries' names extracted from the papers' addresses indirectly represent the authors who wrote the papers. Therefore, one can say where the authors are from, but not individually identify who they are. However, in order to be successful the process needs to incorporate a series of assumptions that are in accordance with the number of authors and addresses indexed per paper. The following section details the devised georeferencing process, and elaborates on the computerized solution developed.

6.2 A Computerized Solution for Georeferencing the GIScience Collaboration Network.

Studies of large data sets that represent a discipline, field or sub-field though their co-authorship networks are not new (Barabasi *et al*, 2002; Wagner and Leydesdorff, 2003; Moody, 2004; Newman, 2004c; Arciniegas and Wood, 2006; Carvalho and Batty, 2006). Unlike them, this study focuses not only on processing a large data set but also, on a georeferencing process that is framed by ISI-WoK limitations regarding authors' addresses (see 6.1.2 section).

Herein, the georeferencing process comprises the time consuming task of extracting authors' addresses and pairing countries from almost 6000 out of the total 12373 scientific documents. Thus, the development of a computerized tool is imperative. The proposed georeferencing procedure is part of the computing tool, detailed in Chapter 3, implemented to build and calculate statistical properties of the GIScience co-authorship network. This section aims to explain in detail the computerized procedure devised to georeference the GIScience co-authorship network that can be applied to georeferencing co-authorship networks built upon the WoK bibliographic format.

6.2.1 The Georeferencing Process

The ultimate goal of the GIScience geo-referencing process is to create a set of 3-tuples in the format (*country i,j collaboration frequency, collaboration country i, collaboration country j)* that geographically represents linkages between co-authors in all included papers. From an algorithmic point of view, the georeferencing process involves the following four steps.

1. A stage that involves reading the data sets and extracting the C1 information and the number of authors for each paper.
2. A step that extracts the co-authors' affiliations (country name) from the C1 tag information.
3. A process that makes geo-referenced pairs using authors' affiliation information (*country from author A, country from author B*).
4. A method that quantifies the strength of collaboration between each pair of countries.

The computational challenge offered by the georeferencing process consists of incorporating the WoK limitations regarding the way that author affiliations are indexed. Thus, step 2 and 3 are the main core of the georeferencing process, leaving one and four as secondary steps at this stage. In general, they involve input and output routines and the performance of some calculations computed from the set of *country, country* pairs. As a Java method part of the main Java tool (see section 3.3.2 chapter 3), the georeferencing procedure uses the *read method* that handles all input and output tasks. The main characteristic of this method is to be able to read a large text file in the ISI-WoK format and to store the papers' information as records comprising of fields that correspond to the ISI-WoK tags. As this method is detailed in chapter 3, this section focuses on elaborating the address extraction and pairing process.

6.2.2 Extracting Author Affiliations

For each paper the georeferencing process is focused on extracting the relevant information concerning the authors' affiliations indexed under the C1 tag. Table 6.40 shows an example of affiliation information under C1. In each paper all authors' affiliations are saved as text strings separated by spaces (see rows in Table 6.40). Here the focus of the attention depends on the aggregation level of the geographical analysis.

Commiss European Communities, Joint Res Ctr, I-21020 Ispra, VA, **Italy**
Commiss European Communities, Joint Res Ctr, I-21020 Ispra, VA, **Italy**
Univ Coll London, Dept Geog, Wetland Res Unit, London WC1H 0AP, **England**
Aristotle Univ Thessaloniki, Sch Agr, Lab Appl Soil Sci, GR-54006 Thessaloniki, **Greece**
Greek Biotope Wetland Ctr, GR-570 Thermi, **Greece**

Table 6.40 C1 Example for an 11-author paper in PERS

For instance, very detailed studies, such as Calero *et al* (2006), concentrated on extracting authors' affiliation at research centre level. Carvalho and Batty (2006) and Batty (2003) focused on postcode (zip code in USA) analysis. While, Wagner and Leydesdorff (2003), Glänzel and Schubert (2004) and Arciniegas and Wood (2006) carried out their analysis at country level. These studies aggregated their bibliographic data by country using the last part of the C1 tag information (between the last "comma" and the end of the string), which corresponds to the countries were the authors were affiliated at the time of the paper (see bold information in Table 6.40).

As this study aims to inspect the geographical distribution of scientific collaboration at the national level, the extracting and pairing process focuses on obtaining and pairing information related to the authors' countries. Thus, for each paper one needs to create a collaboration pairs within the *country, country* format. An intermediate step previous to collaboration pairing is to obtain the authors' countries extracted from the C1 tag. In doing so, the *makingCountryListPerPaper method* from the *UtilityTools Class* extracts the C1 information from each record and returns a string list that corresponds to each one of the countries found. A call to the *extractCountryFromC1* method, belonging to *UtilityTools Class* assembles the countries' list for each paper. The implemented method (Figure 6-40) identifies addresses that do not contain country names but the acronym of any state in the USA and assigns them to USA publications. It also handles compound country names such as Cent Afr Republ, North Korea or Serbia and Montenegro., where necessary.

```
public String extractCountryFromC1(String fullC1)
{
 //C1 is the full address, that needs
 //to be split up into tokens according "\\s"
 String[] fullC1Tk = fullC1.split("\\s");
 int pos = 0;
 int limit=fullC1Tk.length/2;
 String next=null;
 String previous=null;
 String compoundName="";
 boolean coumpound = false;
 //verify that the new lengh is an integer always.
 if (!((fullC1Tk.length%2)==0))
 {
      limit = Math.round(fullC1Tk.length/2);
 }
 //Iterate only until half string length. e.g 10 until 5
 for (int c=fullC1Tk.length-1; c>limit; c--)
 {
 pos = Arrays.binarySearch(countryNames,fullC1Tk[c]);
 if (pos >0)
 {
       if(Arrays.binarySearch(usaStatesAbb,fullC1Tk[c])>=0)
       {
             return "USA";
       }
                       if(fullC1Tk[c].equalsIgnoreCase("ireland")&&fullC1Tk[c-
1].equalsIgnoreCase("north"))
  {
              return (fullC1Tk[c-1]+" "+"ireland");
       }
       return fullC1Tk[c];
      }
      else
      {
       // Not found but I need to find out if it is part of a compound name.
       previous = lookForCountryToken(fullC1Tk[c]);
       if (!(previous==null))
       {
        compoundName = compoundName +" "+previous;
        coumpound =true;
       }
       else if (!(compoundName ==null)&&coumpound)
       {
       return rightOrder(compoundName);
       }
      }
 }
      return "newCountry";
}// End-Method
```

Figure 6-40 ExtractCountryFromC1 Java method, part of the utility java tool developed to handle all operations concerning strings.

The final result is a list that contains a sub-list with all the countries extracted from the affiliation addresses in each paper. This list is read within the method in charge of doing the collaboration pairing operation. However, some irregularities need to be considered regarding the C1 information indexed. The following section explains in details this aspect.

6.2.3 Implications of ISI-WoK limitations within the Geo-referencing Process

Even after extracting the list of countries for each paper, making the combinations is not a straightforward process. Data shows that C1 may comprise one, two or more addresses that may or may not correspond to the same number of authors within each document. For example, for the paper represented in Table 6.39, the country collaboration pair (USA, USA) is a simple case as it observes the same number of authors and addresses. However, the pairing process becomes more complicated for the papers presented in Table 6.40, where the *number of authors = 11* differs from the *number of addresses = 5,* or when there are more addresses than authors.

Cases (for all papers with No. authors > 1)	Proposed Solution	Example (fron any of the 11 journals)	Collaboration Pairing Results
No. authors = No. addresses	Each address correspond to each author, then, make all possible combinations between them.	Langford, M Bell, W Univ Leicester,Dept Geog,Leicester Le1 7rh,Leics,England Cent Int Agr Trop,Cali,Colombia	England, Colombia
No. addresses > No. aut hors	To take the first address and skip the addresses correspoing to the difference, then, make the corresponding combinations.	Bittner, T Winter, S Queens Univ, Dept Comp Sci, Kingston, ON, Canada Queens Univ, Dept Comp Sci, Kingston, ON, Canada Tech Univ Vienna, Dept Geoinformat, A-1060 Vienna, Austria	Canada, Austria
No. authors > No. addresses	To use the existing address to make the combinations.	Abel, DJ Ooi, BC Tan, kc Power, R Yu, JX Csiro,Div Informat Technol,Comp Sci & Informat Technol Bldg,Gpo Box 664,Canberra,Act 2601,Australia Natl Univ Singapore,Dept Informat Syst & Comp Sci,Singapore 0511,Singapore Australian Natl Univ,Dept Comp Sci,Canberra,Act 0200,Australia	Australia, Singapore Austrialia, Australia Singapore, Australia

Table 6.41 List of assumptions made regarding GIScience geo-referencing process.

Consequently, the process of creating *country-country collaboration pairs* requires establishing a set of assumptions that allows the execution from the geo-referencing process of the data source. As listed in Table 6.41, the C1 records can be classified under 3 different cases that help avoiding ambiguity at the time of extracting the affiliation country of all co-authors.

- The **first case** groups documents where the number of *authors equals to the number of addresses.* For 34% out of the number document total, the author pairing is a straightforward process due to the equal number of authors and addresses.

- The **second case** represents 18% of the documents with more addresses than authors. A computerized method scanned all documents showing that in the 90% of these cases the difference between the number of authors and papers is one and, the first and second addresses in the majority of the cases are the same. These two facts suggest that the difference in the number of addresses corresponds mainly to the first author and the proposed solution for this case consists of skipping the addresses corresponding to the difference. Then, using the first and the remaining addresses for making the combinations. In doing so, one assumes that the difference in addresses concerns the first author' past affiliations that has also been indexed. Therefore, losing them does not affect the current author's collaboration links.

- The **third case** collects 48% of the documents which had more authors than addresses. In this case, one assumes that the indexed addresses represent all participating addresses but, they left out the repeated ones. As 44% of the documents in this group exhibit only one address, one assumes that the only address indexed represents a group of authors working at the same organization. The proposed solution is to use the existing addresses to make the combinations.

PT	Journal
AU	Wood, JD; fisher, P; Dykes, J; Unwin, D; Stynes, K
TI	The use of the landscape metaphor in understanding population data
SO	ENVIRONMENT AND PLANNING B-PLANNING & DESIGN
C1	Univ leicester dept geog leicester le1 7rh leics england ; Univ leicester dept geog leicester le1 7rh leics england ; Univ london birkbeck coll dept geog london w1p 1pa england

Table 6.42 Example of a document record with more authors than addresses.

Table 6.42 shows an example of this last type. At the time of writing the papers in 1999, JD. Wood, FP. Fisher and J. Dykes were based on Leicester University, while D. Unwin and K. Stynes were affiliated to Birkbeck College, University of London. Thus, it seems very likely that in this type of cases, all addresses are represented and indexed, but they are not repeated where there is more than one author at the same institution. Here, one assumes that for the 56% of documents with more authors than addresses, the missing collaborations correspond to the strength of the link between authors in the same country.

6.2.4 Producing the Country Collaboration Output

```
public void makeAuthrosCollaborationsCombinations()
{
 // How many one authors and c1 addresses per paper?
 int howmanyEqual=0, inTotal =0;
 //equal addresses and authors
 //For making the combinations
 for (int i = 0; i < authorsPapers.size();i++)
 {
  // START DOING THE COMBINATIONS.
  if (numAuthors==numCountries)
  //1. no. Authors = Nno. C1 addresses so 1-1 relation
  {
  //Do the pairing because is 1-1 relationship
      for (int j=0; j< numCountries-1; j++)
      {
       String country1 = c1CountryTk[j];
       For ( int k=j+1; k< numCountries;k++)
       {
       String country2 = (c1CountryTk[k]);
       pairedAuthorsAddresses.add (country1+"\t"+country2+"\t"+currentYear);
       }
  }
  ........
      }
       else if(numCountries>numAuthors)
  //2. If there are more addresses than authors and the diff is one.
       {
       .....
       int difference = numCountries-numAuthors;
       for (int k =difference+1; k< numCountries;k++)
       {
        pairedAuthorsAddresses.add
   (c1CountryTk[0]+"\t"+c1CountryTk[k]+"\t"+currentYear);
       }
       //To make the remaining combinations.
       for (int l = difference+1; l< numCountries-1;l++)
       {
   ......
             for (int t= l+1; t< numCountries;t++)
             {
              \\ Add Combination In the Format, Country1,Country2, Year
   pairedAuthorsAddresses.add
   (c1CountryTk[l]+"\t"+c1CountryTk[t]+"\t"+currentYear);
   }
       }
       ......
       }
       else //3. If there are more authors than addresses.
       {
       String country1=null;
       if (numCountries==1)
   // 3.1 If there one addresses, Assign it to all combinations.
       {
             country1 = c1CountryTk[0];
             for (int j=0; j< numAuthors-1; j++)
   {
   for (int k=j+1; k< numAuthors;k++)
              {
              pairedAuthorsAddresses.add
    (country1+"\t"+country1+"\t"+currentYear);
   \\Add Combination In the Format, Country1, Country2, Year
              }
        }
       ........
        }
```

```
        else
  // 3.2To the remaining cases, make the combinations among the
  // existing addresses.
        {
         for (int j=0; j< numCountries-1; j++)
         {
                       ......
              country1 = c1CountryTk[j];
              for (int k=j+1; k< numCountries;k++)
              {
              String country2 = (c1CountryTk[k]);
              pairedAuthorsAddresses.add
       (country1+"\t"+country2+"\t"+currentYear);
    \\Add Combination In the Format, Country1, Country2, Year
              }
             }
      .........
        }
  }// End-Most External For
} //End-Method
```

Figure 6-41 *AuthorsCollaborationCombination* Java method generates *country, country, year* combinations for all papers. Only the main instructions are displayed.

The method uses the list of countries assembled by *ExtractCountryFromC1 Method.* Thus, a series of nested if-statements (see Figure 6-41) are used to select which one of the cases (explained in the previous section) can be applied to extract the countries from the addresses of the current paper. Accordingly, for each paper a list of 3-tuples in the format *country1, country2, year* is assembled. Each 3-tuple represents a scientific collaboration from author1 located in *country1* with author2 located in *country2* made in this specific year. For each paper, the same process is repeated as many times as defined by the binomial coefficient (defined in Chapter 3), depending on the number of co-authors. The intermediate result is a new list that contains all 3-tuples for all papers in the data set. Finally, a third method named *printCountryTuplasFrequencies* takes the previous list adding up all equal 3-tuples generating the frequency of the scientific collaboration between *country1, country2* in that specific year. This list is the ultimate result from the georeferencing process that will be used to create the maps that geographically characterises the scientific collaborations in GIScience from the years 1992-2002.

6.3 Spatial Characterization

Generating the authors and country collaboration 3-tuples enables the exploration of the collaboration links between countries and to geographically characterize the part of the GIScience academic network that the co-authorship network studied represents. Two types of collaboration linkages were generated (see section 6.2), one that represents collaboration at author level using their countries as georeferencing points and the other that represents international collaboration between countries.

6.3.1 The Geographical Distribution of Authors in the GIScience Collaboration Network

First, scientific links between all co-authors within a paper were created, using their affiliation countries as georeferencing points. At the same, one will be able to explore the geographical distribution of co-authorship in the GIScience network[9]. It implies that the linking process for a given paper does not distinguish between local (same country authors) or international collaboration (between authors from different countries). Instead, it creates all possible links between all co-authors' affiliation countries. The number of times the combinations is calculated according to the binomial coefficient (BC), depending on the number of co-authors a paper has. For example, if a paper is written by 10 authors from only one country, following the BC formula, there would be 45 collaboration pairs for that country.

$$\text{BC} = \binom{n}{k} = \binom{10}{2} = 45$$

Where *n*, number of co-authors = 10 and *k*, elements in the combination = 2. In general, the total author-collaboration links per country indicates the number of times an author from that specific country has participated as a co-author within the scientific works, which are part of the network. In order to study changes of co-authorship patterns between 1992-2002, the total number of collaboration links was cumulated per country, and divided into four time periods starting in 1992. Though, the 1992 results will be influenced by the short time period covered, these figures will give an insight in the network connectivity at the beginning of the study period. The second time interval corresponds to the first part of the nineties (1992-1996), the third one to the second part of the nineties and the beginning of next decade (1997-2002) and finally, an interval that covers the full time period, (1992-2002).

In fact, 51% of the author-collaborations, out of the 26808 established between 1992-2002, involved only 3 countries: USA, England and Canada (for the full list see Appendix G)[10].

[9] Within the context of this chapter GIScience Co-authorship Network refers specifically to the 5-4 GIScience collaboration network.

[10] It is important to note that for this chapter the United Kingdom (UK) is as separated into England, Northern Ireland, Scotland and Wales. There are two reasons for this, papers' addressees present these countries separated, and to avoid inflating figures by scholars located in England. Batty (2003) revealed a geographical concentration of highly cited authors around central London area.

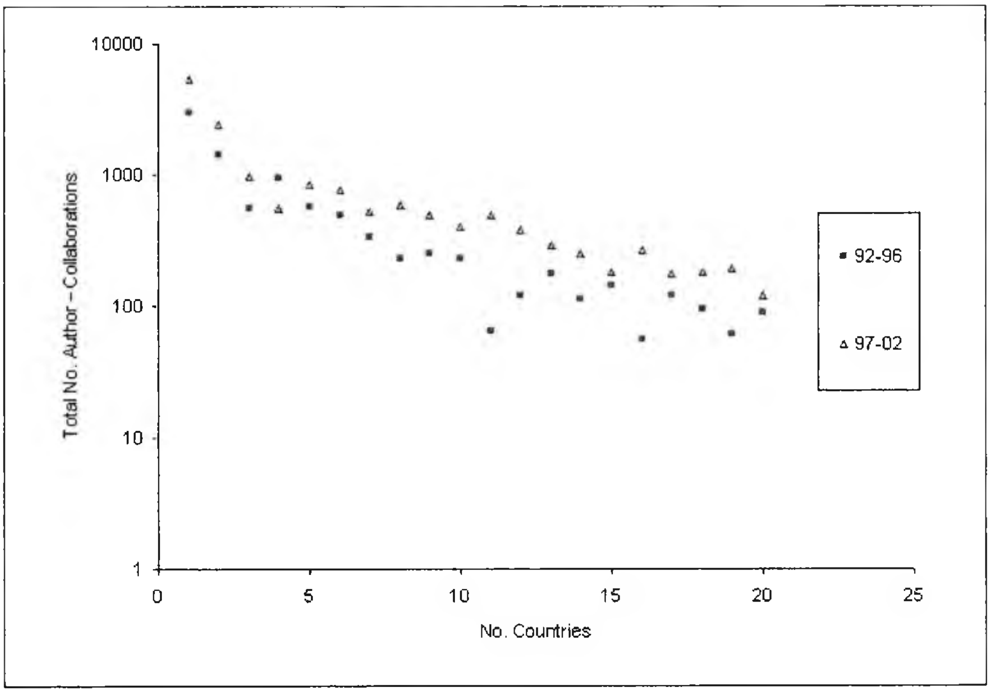


Figure 6-42 Log-Plot showing the frequency distribution of the total number of author-collaboration per country during 1992-1996 and 1997-2002.

Table 6.43 illustrates how the participation from USA, England and Canada, all English-speaking countries, have doubled between the first and the second part of the decade.

	1992-1996		1997-2002		Increment or Decrement Ratio	
Country	International papers	Authors participation	International papers	Authors participation	International papers	Authors participation
United States	205	2960	403	5423	2	2
England	160	1417	318	2445	2	2
India	16	948	25	556	2	1
France	53	563	101	834	2	1
Canada	68	560	143	969	2	2
Italy	45	489	88	775	2	2
Australia	37	332	67	527	2	2
Netherlands	36	249	87	494	2	2
Germany	43	227	97	585	2	3
Spain	21	227	47	402	2	2
Scotland	48	174	91	293	2	2
Greece	26	142	30	179	1	1
Japan	10	121	48	376	5	3
Wales	38	118	64	174	2	1
Israel	19	114	33	247	2	2
Russia	19	105	20	85	1	1
New Zealand	9	95	36	180	4	2
Sweden	10	89	25	119	3	1
China	12	64	66	492	6	8
Brazil	11	61	51	190	5	3

Table 6.43 Top 20 Countries with the highest author-participation and their total number of international collaborative papers between 1992-1996 and 1997-2002. The far right two columns show the increment or decrement radio regarding the country's author-participations and number of international papers.

In the second part of the decade there are 8 times more Chinese authors co-authoring GIScience papers, than if compared to the numbers from the first half of the decade. Moreover, China increased its participation in international papers that include authors from at least one other country. However, comparing China's international participation to that of the USA or England which doubled their contribution, it is important to take into account the amount of human resources involved in moving participation from 150 to 300 (English increment) , compared with 12 to 66 papers (Chinese increment).

Individual statistics on author-collaborations by country do not communicate much about the links betweens countries per se. Hence, using the graph paradigm (explained in detail in Chapter 3 and 4) and Pajek software, four author collaborations networks, one for each time period, were built.

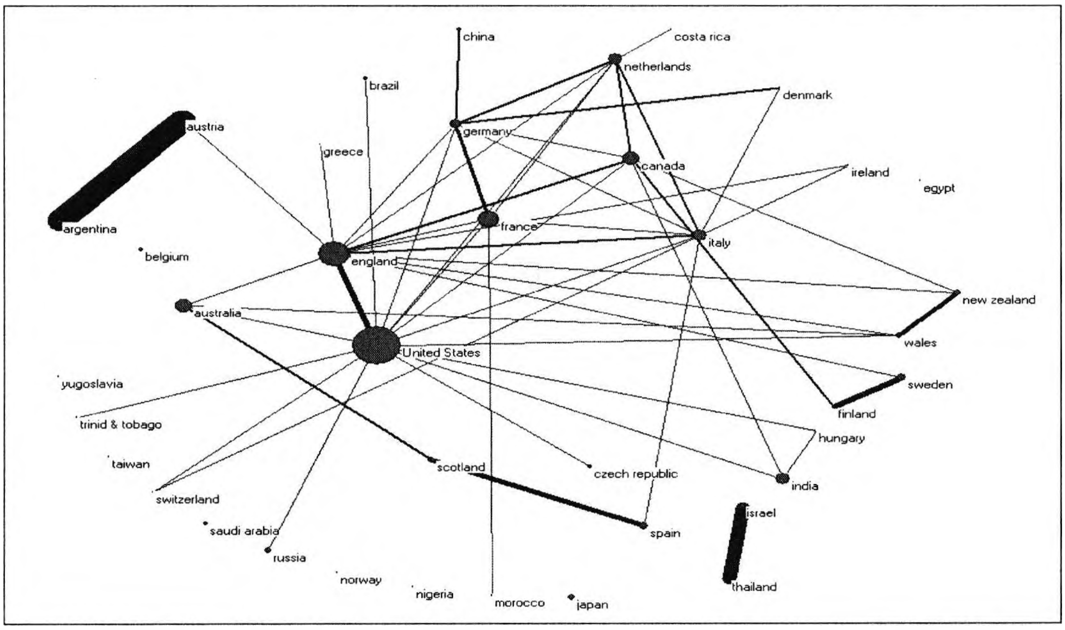


Figure 6-43 GIScience author-collaboration network by country in 1992. The network maps the initial state of international collaborations around GIScience topic within the context of this study.

Within the author-collaboration network in Figure 6-43, nodes represent countries linked by international collaboration that involved authors from different countries. The size of the nodes is proportional to the total number of times authors from the given country have internationally participated as co-authors. The network in Figure 6-43 shows the disproportionate presence of author-collaborations involving English-speaking countries such as USA, England, Canada or Australia; accompanied by a less stronger, but still noticeable, presence of authors from European countries such as France, Italy or the Netherlands. The Collaboration Exclusivity Index (CEI) defined by

Arciniegas and Wood (2006) is used to establish the strength of the link between two countries (nodes). The CEI is defined as:

$$CEI_{ij} = \frac{n_{ij}}{n_i + n_j}$$

Equation 6-11 Collaboration Exclusivity Index (CEI)

Where n_{ij} is the number of papers with co-authors from countries *i* and *j*, n_i is the total number of papers from country *j* and n_i is the total number of papers from country *i*. The CEI represents collaboration in terms of "how exclusive" is the relationship between the two countries. In that way, the higher the CEI, the more exclusive the collaboration between the two countries, then, the thicker the line linking the two in the network graph. Observing the distribution of links, Figure 6-43 shows that authors from USA and England have a diversified set of collaboration partners, but, the USA-England relationship is stronger than the others (England-USA CEI = 4%), compared with more sporadic collaboration with colleagues from other countries (relationship showing smaller CEI figures). Collaborations between European countries also show strong CEI between them. At the other extreme, the sporadic participation of Argentinean authors are mainly with Austrian colleges thus, the Argentinean-Austrian relationship is very strong.

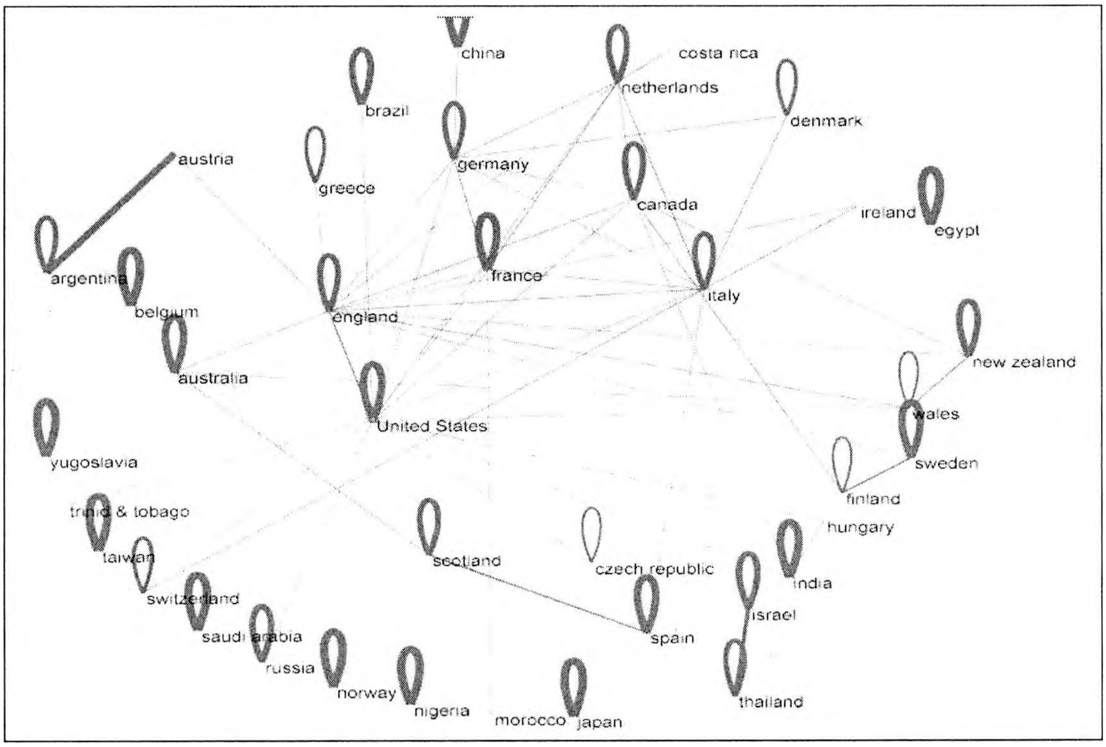


Figure 6-44 The strength of GIScience national and international author-collaborations during 1992. The thickness of the links between countries is proportional to the CEI defined by Equation 6-2 respectively.

Figure 6-44 reveals that in most cases, the strongest and most exclusive relationships are local, rather than international scientific collaborations. Therefore, not all of one country's multi-authored papers are the product of international collaboration; many of them are national works. The most exclusive relationships (thicker links) for the most participative countries (with the higher number of authors co-authoring the selected papers) such as USA, England or Canada are between local authors. The same pattern is observed in low participant countries such as Egypt, Taiwan or Saudi Arabian where GIScience work is the product of local collaborations.

Figure 6-43 and Figure 6-44 show author participations at country level during the network starting point. To be able to identify steadier trends or collaboration patterns (if any), the same procedure was applied to cumulate and plot data from 1992-1996, 1997-2002 and 1992-2002. However, in order to separate one-off or sporadic from more frequent collaboration linkages, and to expose truly trends, two thresholds were set up. In that way, only relationships in which both countries have an author-participation greater than 5 times and with a CEI of at least 1%, were mapped. In doing so, one will be able to see if some of the co-authorship linkages that started in 1992 grew stronger, or did not even pass the threshold that made them disappear from the subsequent maps.

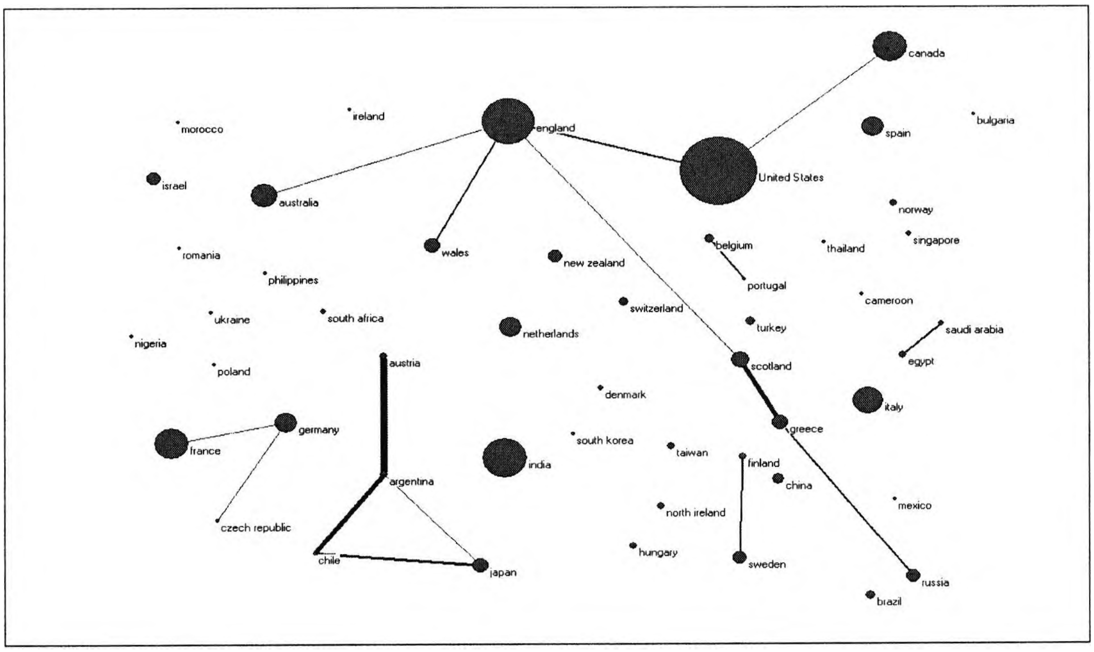


Figure 6-45 1992-1996 GIScience authors collaboration network, displaying country pairs with CEI > 1% and each country having at least 5 author-collaborations in total.

Within 1992-1996 period, the collaboration network in Figure 6-45 shows that the stronger and most exclusive collaboration linkages are among English-speaking countries. The thresholds filtering weaker links, reveals that apart from England, the other most frequent collaborations from USA authors involved Canadian researchers. However, in the case of England, the graph displays a much wider set of collaborations, unsurprisingly with other UK countries (Wales and Scotland) and Australian authors. Also the 1992-1996 network shows the shaping of other sub-networks such as Nordic (Sweden and Finland) and European (German and France) countries.

Moving to 1997-2002, the author-collaboration network in Figure 6-46 reveals a much stronger predominance of English speaker countries. However, it is important to note that the USA-England partnership becomes less exclusive, as its CEI decreases from 2.1% in 1992-1996 to 1.3%.

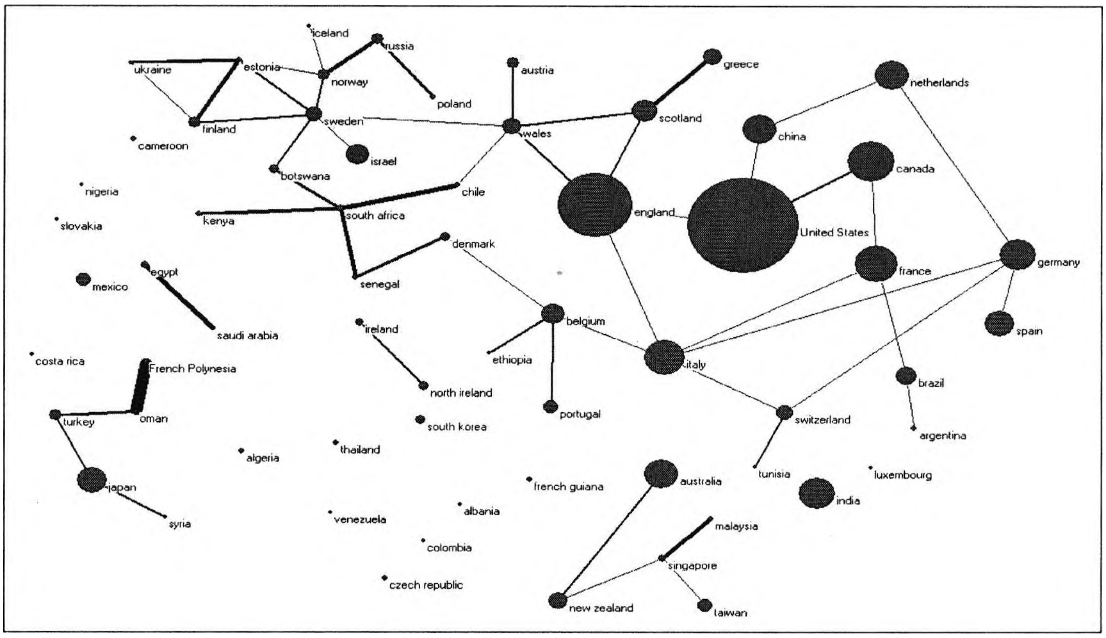


Figure 6-46 1997-2002 GIScience author-collaboration network, showing country pairs with CEI > 1% and with both countries having at least 5 author-collaborations in total.

In that sense, the 1997-2002 network shows that authors from these countries have a more diversified group of co-authors. In the case of England, the obvious strong collaborations between UK countries prevailed, and in the case of the USA a strong relationship with Chinese authors appeared. In contrast to 1992-1996, during 1997-2002 European collaborations became stronger, as links between the Netherlands, Germany, Spain, Italy and France grew. The network also shows that collaborations were not sporadic between Nordic countries such as Sweden, Norway, Iceland and Finland in the previous period. Some Nordic countries also extended their relationships with ex Soviet Countries such as Estonia, Ukraine and Russia. Other sub-networks become apparent comprising countries geographically close such as in Africa like South Africa, Botswana, Kenya and Senegal or from Far East and Oceania such as Singapore, Malaysia, Taiwan, Australia or New Zealand, grew stronger.

The co-authorship network covering the full time period, 1992-2002 (Figure 6-47), visualizes the significant difference in author participations from English speaking and European countries, compared with those from other countries.

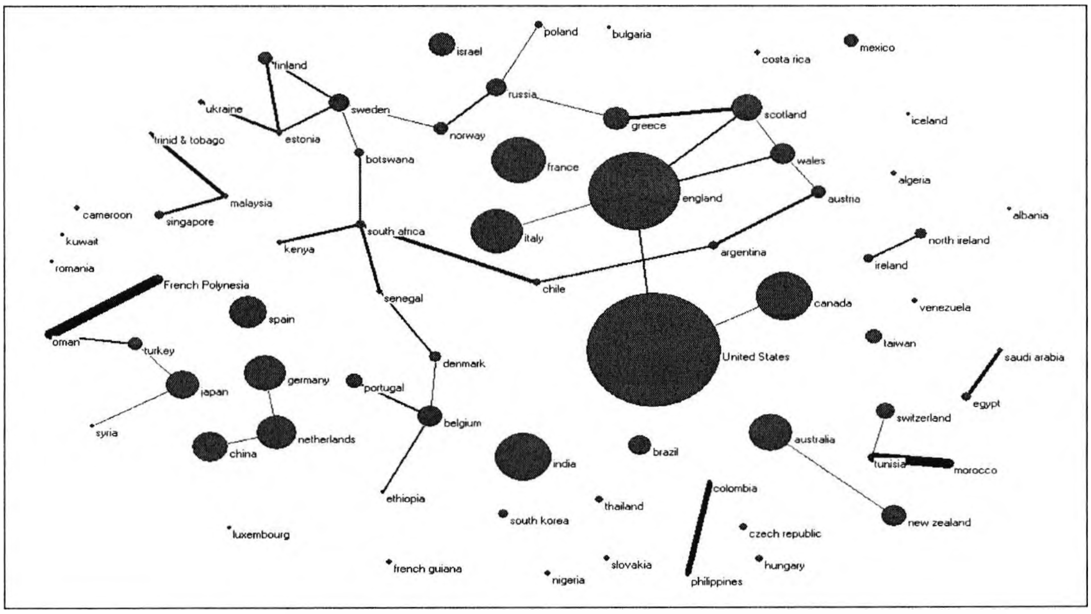


Figure 6-47 1992-2002 GIScience author-collaboration network, showing country pairs with CEI > 1% and with both countries having at least 5 author-collaborations in total.

It shows that English and Canadian authors are the most regular and exclusive partners to USA researches, as shown in all previous networks. A similar situation is observed for Australian with New Zealander, German and Dutch authors, and also between Nordic countries such as Finland, Sweden and Norway.

Nevertheless, as shown in Figure 6-48, the strongest and most predominant co-authorship linkages within the analyzed scientific literature are from authors in the same country. For example, CEI of English authors involves 20.4% of the links during 1992-1996, 21.9% during 1997-2002, and they represent around 19.51% of the links for the 1992-2002 overall network. Similarly, USA-USA links represent 22.2% during 1992-2002, 21.5% during 1997-2002, and 21.7% during the overall time. This is the pattern for the majority of author-collaborations extracted from the studied GIScience literature.

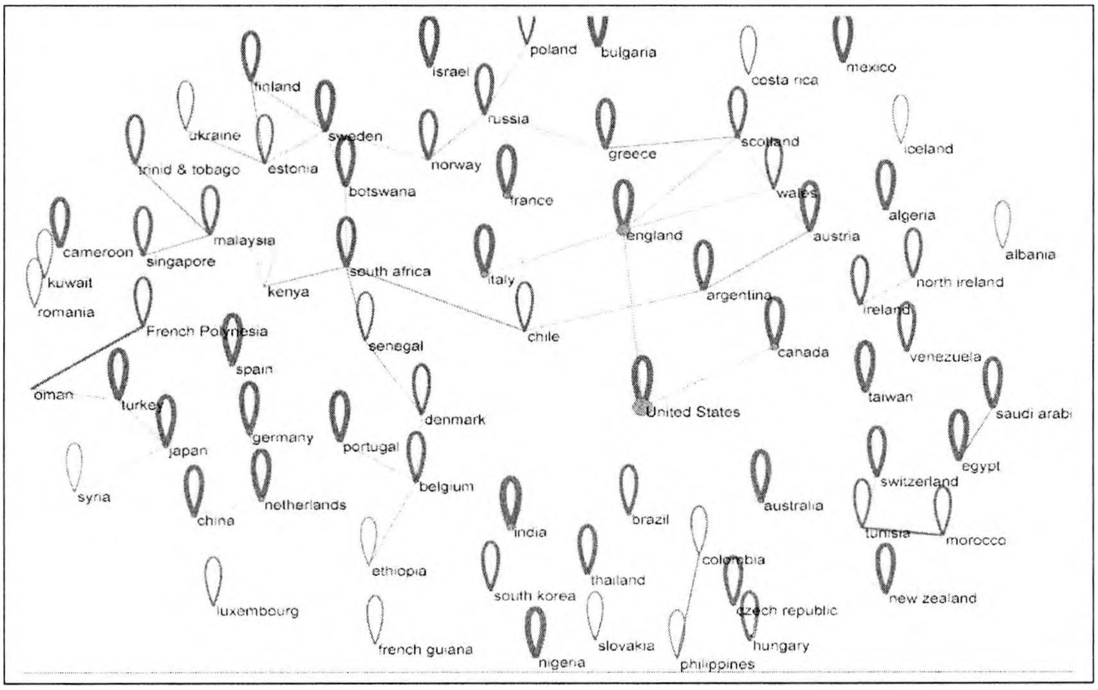


Figure 6-48 1992-2002 GIScience author-collaboration network. The sizes of the nodes and links between countries are proportional to the total number of author-collaborations per country and to the CEI defined by Equation 6-2 respectively.

In order to explore truly international patterns between countries, the following section concentrates on analyzing international country collaborations, excluding those collaborations with authors from the same country.

6.3.2 Cross National Collaboration on GIScience

This section focuses on exploring the geographical distribution of the international collaboration linkages gaining insights about cross-national publication patterns on GIScience. In doing so, one will focus on collaborations between scientists affiliated in different countries, referred to by Glänzel and Schubert (2004) as “cross-national level”. As a consequence, within the study of cross-national collaborations only linkages between authors in different countries are taken into account, excluding papers co-authored by authors in the same country. In other words, if there is a paper in which there is more than one co-author per country, then, the country is only used once to make the respective collaboration pairs. In doing so, totals per country reflect the number of papers in which authors from a given country have international collaboration.

6.3.3 Mapping the Global GIScience Co-authorship Network

The GIScience collaboration network is made up from collaborations extracted from 12373 papers. However, it is important to note that only around 52% of these papers are collaborative works (by at least 2 co-authors). Among collaborative works, only papers that include at least two different national affiliations were selected. Taking into account these restrictions, the share of international collaboration papers for all time periods shows that English-speaking countries such as USA, Canada and England account for the majority of the international participations.

The data in Table 6.43 shows that the international participation of the USA doubles from 205 papers during 1992-1996 to 403 papers during 1997-2002. Similarly, English participation increases from 160 papers during 1992-1996 to 318 papers during 1997-2002. However, individual statistics by county do not inform on how the structure of the GIScience collaboration network at cross-national level has changed over the time under study. In order to visualize the network's structural changes, the links of the most active countries were analyzed the years 1992-1996 and 1997-2002. Similar to Glänzel and Schubert (2004), the Salton's measure (defined in Equation 6-1) is used as an indicator of international collaboration strength. Moreover, to differentiate frequent from sporadic international collaborations, only country pairs with SM > 1% and when each country has more than 5 international papers are selected.

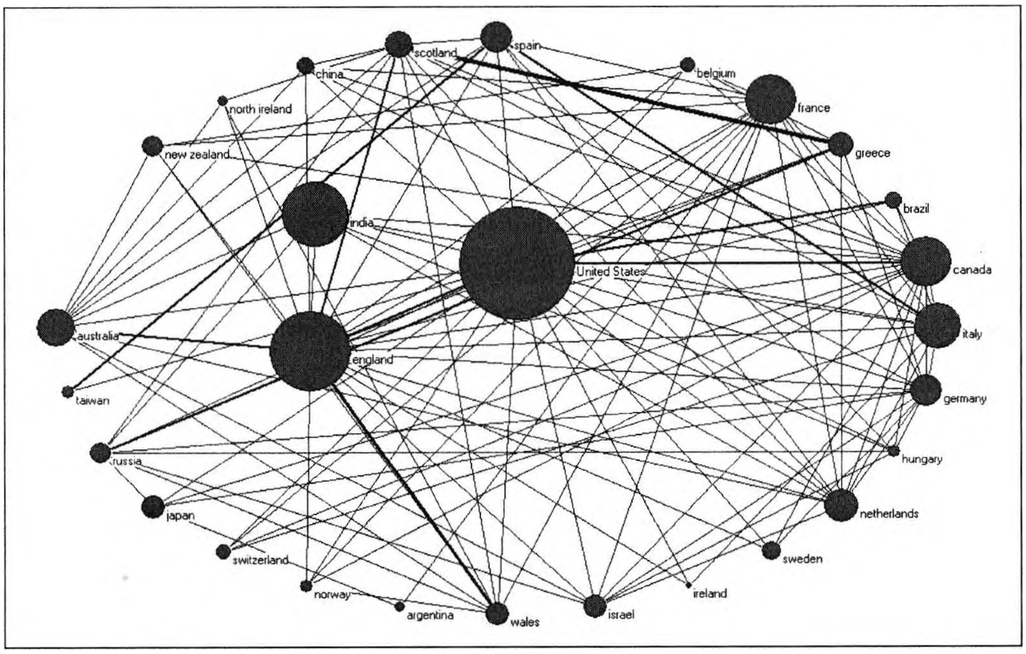


Figure 6-49 1992-1996 Global GIScience country-collaboration network, displaying country pairs with SM > 1% and countries with at least 5 participations.

The previous author-collaboration data shows that the most exclusive relationships were between authors from the same country. However, removing same country collaboration allows the analysis of the distribution of international links for a given country.

During 1992-1996, the global collaboration network in GIScience (see Figure 6-49), reveals a strong presence of two of the three "hub countries", USA and England, detected within the country participation network for the same time period (see Figure 6-45); however, India also becomes a strong player. It is not only in GIScience that USA and England play an important role, Glänzel and Schubert (2004) found similar patterns mapping the co-authorship network of global science. Not only are researchers from these two countries hubs within the global collaboration network, but also their research is the most cited, as Batty (2003) revealed.

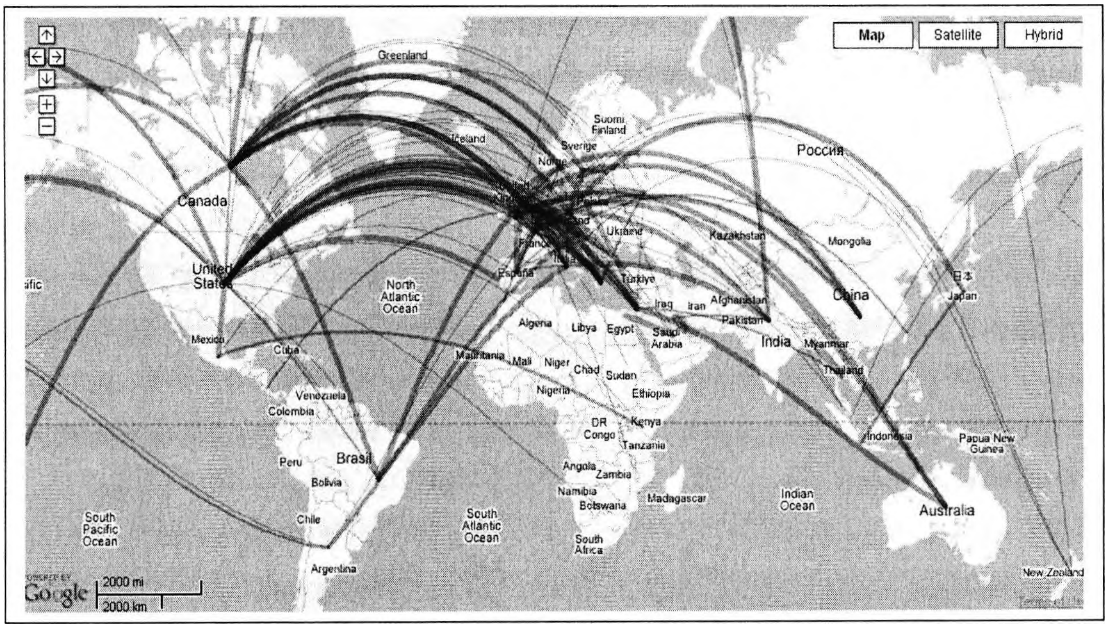


Figure 6-50 Geographical distribution of GIScience country-collaborations from 1992-1996, displaying county pairs with SM>1% and countries with at least 5 participations.

Within the network strong bilateral linkages between the USA and England predominates, and asymmetric relationships between these two countries and all other countries. Apart from USA-England relation, bilateral relationships are identified between European countries that have assembled a large collaboration cluster (see Figure 6-50). According to SM, the most exclusive international partnerships (see Appendix H for the full country collaboration list) are between Greece and Scotland. Their scientific links account for the 42% of total links of each country, followed by England and Wales with 32%, USA and England with 20% and USA and Canada with 19%.

Mapping the collaboration network, the spatial distribution suggests common preferences within stronger partnerships such as language and cultural affinities, geographical location and administrative divisions. Excluding England links, a zoom in at the European countries (Figure 6-51) shows how actively symmetric collaboration took place between 1992-1996 among them. It reveals that strong partnerships were forged within the European countries such as Spain, France, Italy and the Netherlands referring to GIScience.

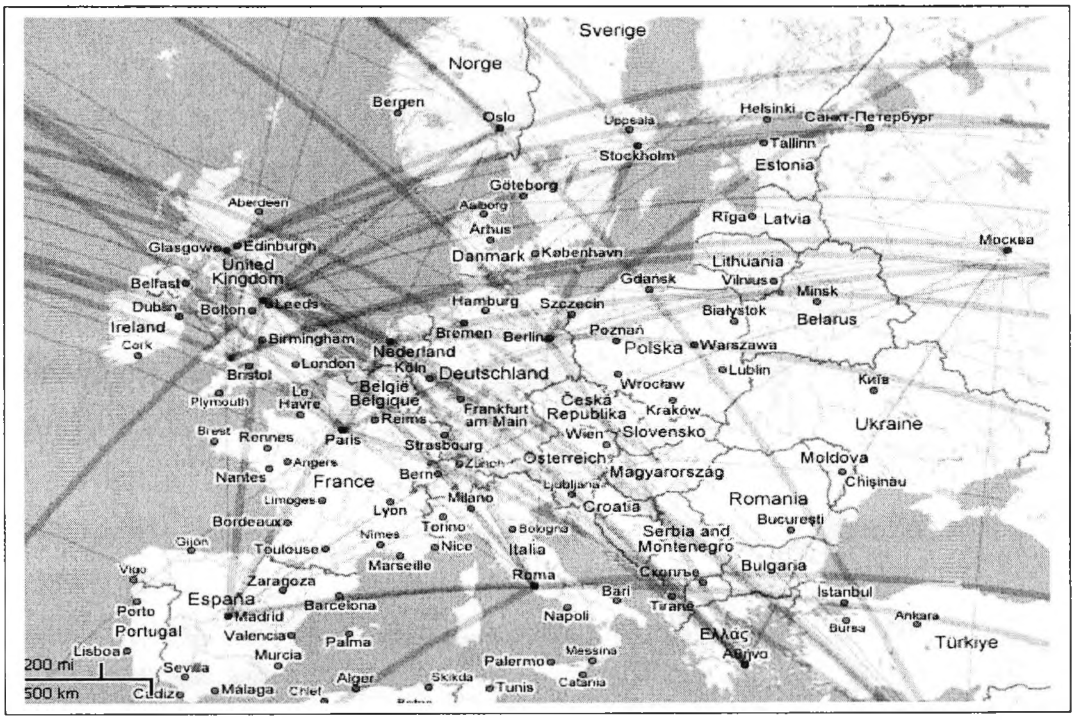


Figure 6-51 Zoom in the Geographical distribution European GIScience country collaborations map from 1992-1996, displaying countries with at least 5 international participation and county pairs with SM>1%.

The 1997-2002 co-authorship country network in Figure 6-52 reveals more intensified international collaboration in GIScience compared to the previous period; despite both networks being built from around the same number of papers. One can say that papers from 1997-2002 where written by more international set of co-authors, resulting from more international links between countries.

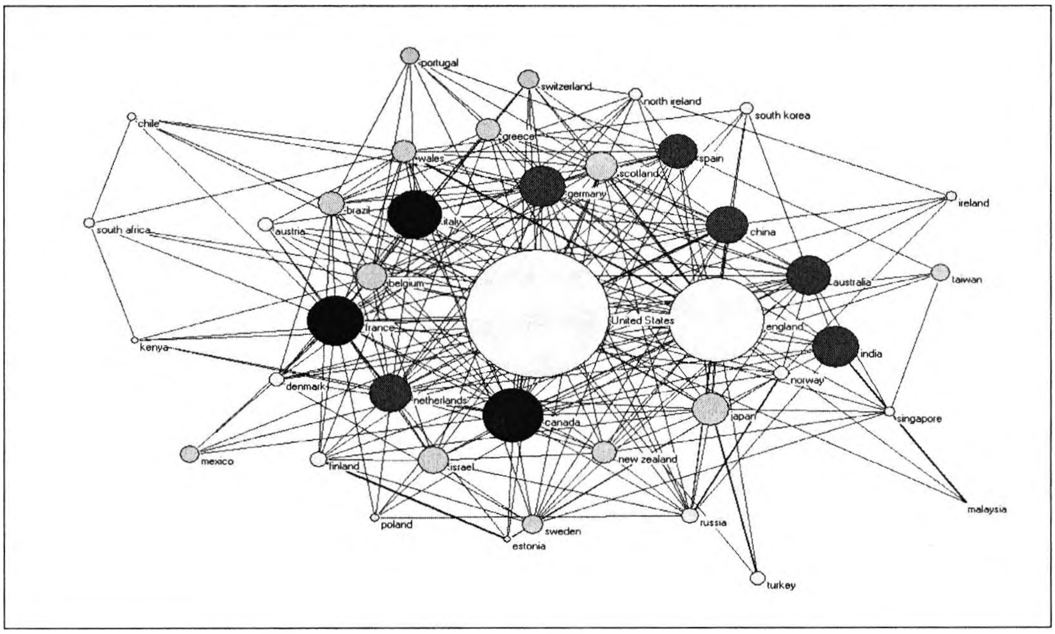


Figure 6-52 1997-2002 Global giscience country collaboration network, displaying countries with at least 5 participations and with SM>1%. Countries are classified into 5 categories depending on the total number of paper collaborations. Yellow nodes between [20-99], light green between [100-400], red between [401-600], blue between [601-1000] and finally, pale pink nodes with more than 1000 collaborations.

For the overall period, as the differences on collaboration numbers become more accentuated, five groups of countries can be easily identified One corresponding to the major international collaboration players (pale pink) such as USA and England, the second (blue) comprises of very active collaborators France, Italy and Canada. A third group of intermediate players (red nodes) such as India, Australia, Spain, China, Germany and the Netherlands. A Fourth group (light green) comprises secondary collaborators such as New Zealand, Switzerland, Japan, Israel, Greece, Brazil or Belgium, and a fifth group (yellow nodes) with the least international collaborator countries such as Austria, Chile, Denmark or Estonia.

Comparing the SM data from 1992-1996 and 1997-2002 time periods, one is able to detect whether or not country1-coutnry2 partnership becomes stronger, died out or, if new relationships were forged. Table 6.44 lists the lost partnerships from one to another period. The table suggests that partnerships such as Japan-Argentina, Northern Ireland –Scotland or Germany-Hungary, which used to represent more than 10% of the total international collaboration links of the respective countries during 1992-1996, did not have continuity during the second time period. Probably, some collaborations were made between these countries during 1997-2002, but they were less exclusive (SM < 1%). It may be due to the fact that the countries diversified their international partners more.

Country 1	Country 2	SM
Japan	Argentina	13%
Northern Ireland	Scotland	12%
France	Argentina	11%
Germany	Hungary	10%
Sweden	United States	9%
Hungary	India	8%
Canada	Hungary	8%
Israel	Hungary	8%
Russia	Hungary	8%
Spain	Hungary	7%
Australia	Northern Ireland	7%
Northern Ireland	Wales	7%
Wales	Norway	7%
Italy	Scotland	6%
Italy	Ireland	6%
New Zealand	Wales	5%
Sweden	Germany	5%
Brazil	Italy	4%
Canada	Sweden	4%
Israel	Australia	4%
India	Italy	4%
Wales	Russia	4%
Australia	Spain	4%
England	Brazil	2%
United States	Hungary	2%

Table 6.44 Lost of international collaboration partnerships (and their SM indexes) forged during 1992-1996, but with no relevant presence during 1997-2002.

Conversely, some collaboration partnerships became stronger from one period to another period (see Table 6.45). Not surprisingly, the USA-China link grew fast, from 4% in early nineties to almost 17% during 1997-2002, representing an increase of around 13%.

country1	country2	1992-1996 SM	1997-2002 SM	SM Increment
United States	China	4%	17%	13.1%
Wales	Scotland	2%	9%	6.8%
England	Scotland	18%	24%	5.8%
Canada	France	2%	7%	5.8%
Northern Ireland	England	6%	11%	4.8%
Australia	New Zealand	5%	10%	4.7%
Japan	United States	9%	12%	3.4%
Netherlands	Germany	8%	11%	3.3%
Germany	United States	6%	10%	3.2%
India	United States	7%	10%	3.0%

Table 6.45 Top 10 collaboration partnerships (the highest SM indexes) forged during 1992-1996 that strengthened during 1997-2002.

It suggests that from the earlier period, China has become an important partner for USA scientists. The same growing pattern is observed for USA-India collaborations, but at a slower pace of growth compared with the Chinese one. Also, intra UK countries, New Zealand and Australia and the Netherlands and Germany became closer partners during the second part of the nineties and early two thousand.

Country1	Country2	1997-2002 SM
Malaysia	Singapore	22%
Estonia	Finland	17%
Singapore	India	16%
Japan	Turkey	15%
Norway	Russia	15%
Chile	South Africa	14%
Finland	Sweden	14%
Kenya	South Africa	11%
Germany	Switzerland	10%
New Zealand	United States	9%
Belgium	Italy	9%
Belgium	Portugal	9%
Norway	Sweden	9%
Netherlands	China	8%
Russia	Poland	8%
Estonia	Norway	8%
Finland	Italy	8%

Table 6.46 Top 10 New More Excusive Collaboration Partnerships (and their SM indexes) forged during 1997-2002 period.

Moreover, a new set of collaboration partnerships arose during 1997-2002 (see Table 6.46). Among them, the Malaysia-Singapore partnership shows the strongest, representing 22% of the total collaboration links of both countries. Between the most exclusive new relations, it is important to note the increase of the participation of Nordic countries (Finland, Estonia, Sweden and Norway) by establishing new connections between them, or through new partnerships such as with Russian or Italian authors. However, England followed closely by USA are the countries with the highest number of new collaboration partnerships. Moreover, SM values shows that the new English relations are stronger that those forged by USA. For England, the most exclusive new relation is with Switzerland (SM =7%), and for USA the stronger ones involves authors from Turkey, South Africa and Portugal, all with around SM =2%.

The figures in Table 6.47 show an interestingly increase of activity around GIScience in developing countries such as Brazil, India and China. Also, it is important to note the increasing collaboration pattern exhibited by European countries such as Germany, France, Sweden, Denmark, Spain and Italy. In doing so, they appear to be expanding their scientific collaboration frontiers.

Country	Total New Collaboration Partnerships	Country	Total New Collaboration Partnerships
England	17	Portugal	8
United States	16	Greece	7
Belgium	14	New Zealand	7
Brazil	14	Scotland	7
Germany	14	Chile	6
Canada	12	Estonia	6
France	12	Kenya	6
Japan	12	Norway	6
Sweden	12	Poland	6
China	11	Russia	6
Denmark	11	South Africa	6
Wales	11	South Korea	6
Australia	10	Switzerland	6
Finland	10	Israel	5
Netherlands	10	Mexico	5
Austria	9	India	4
Singapore	9	Ireland	3
Spain	9	Malaysia	3
Italy	8	Taiwan	3
Northern Ireland	8	Turkey	3

Table 6.47 Complete list of total numer of new international collaborations during 1997-2002

Mapping GIScience country collaboration during 1997-2002 allows the visualisation of all new partnerships forged during this period around the GIScience topics.

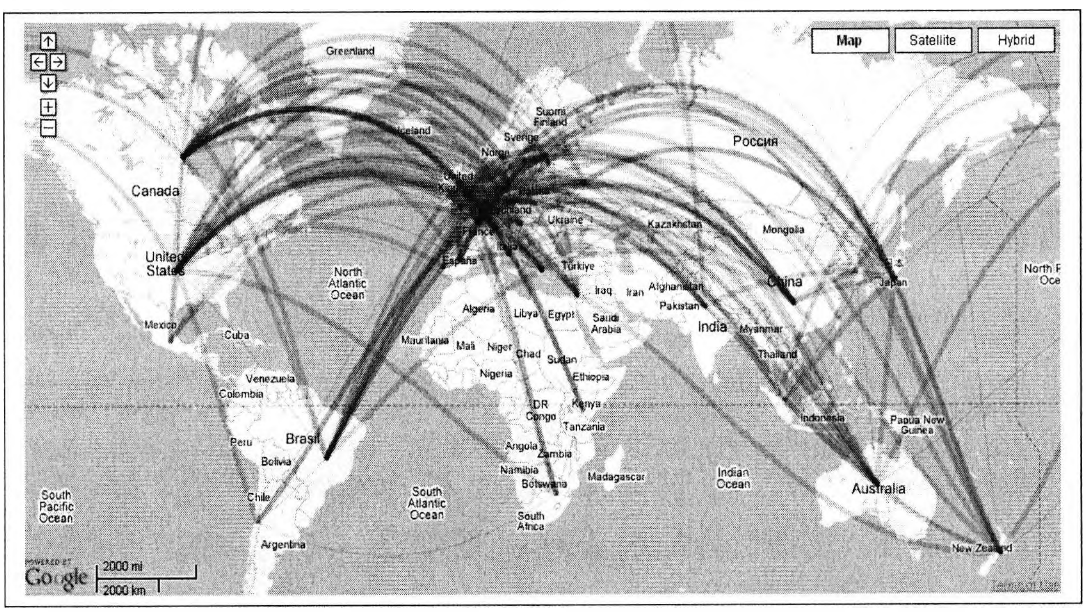


Figure 6-53 Geographical Distribution of 1997-2002 international GIScience country Collaborations, displaying countries with at least 5 participations and county pairs with SM>1%

While comparing the 1997-2002 global map in Figure 6-53 and the 1992-1996 global map (in Figure 6-50), one can notice the increase of international collaboration of countries such as Brazil, Chile, South Africa or Kenya. The map shows how USA and Canada have intensified their existing collaboration links and have forged a new set of geographically dispersed partnerships.

6.3.4 Globalization and Internationalization in GIScience

The revealed patterns bear out general perceptions of scientific outputs, at the global level Glänzel and Schubert (2004) descried USA, England and Canada as very active countries in all scientific fields. Moreover, exploring journal internationalization in human geography, a related field to GIScience, Gutierrez and Lopez-Nieva (2001) identified USA, UK, Canada and Australia as the countries with the greatest scientific output in the journals they selected[11]. Similarly, this study identified USA and England as countries with a disproportionate amount of scientific participation (in terms of number of authors) and scientific output in the GIScience collaboration network. Additionally, figures reveal the rising of new players such as India and China, and also identified an increase of collaboration involving European countries, not only those inside the European Union (EU), but also with partners all over the world.

As the collaboration network is biased toward papers written in English, lingual preferences cannot be fairly assessed. However, one can note the strongest and highly exclusive and asymmetric partnerships between English-speaking countries such as USA, Canada, England, Scotland, New Zeeland, Australia, Northern Ireland and Wales. Besides, strong collaboration patterns between African or Nordic Countries, countries belong to the EU and Australia and New Zeeland may suggest geographical and political preferences at the time to collaborate. However, none of these relationships show an exclusivity level (SM) that indicates a tendency to only collaborate with those countries. In contrast, figures show that even though some countries exhibit frequent collaborators, they also show a tendency to expand their collaboration frontiers.

[11] It is not strange that Journal of Geographical Information Science, Progress in Human Geography and Environment and Planning are included in the Gutierrez and Lopez-Nieva (2001) and in the present study's data sources, as Geography is an essential player in both fields.

7 Conclusions and Further Work

The overall aim of this thesis has been to build co-authorship networks that allow exploration and better representation of the multidisciplinary nature and fuzzy disciplinary borders of GIScience. In doing so, this study has argued that *i*) GIScience's theoretical framework and research findings are relevant to a wide variety of disciplines including traditional geosciences and environmental fields, but also, other fields where geographical location is relevant, *ii*) GIScience creates a network of scientific collaborations that can be interpreted as the interplay between works published on core and peripheral journals to the discipline.

This closing chapter evaluates the success of the proposed approach by re-assessing each one of the stated research objectives, discusses disadvantages and possible further work.

7.1 Re-assessing the Aims and Objectives

It has been the aim of this study to build collaboration networks for representing and exploring the multidisciplinary nature and fuzzy disciplinary boundaries of GIScience. This study argued that it can be done through *i) the* analysis of the discipline co-authorship network structures *ii)* the extracted from bibliographic repositories with a core and peripheral journals and *iii)* using the variation in the concentration of scientific outputs in the publication forums to measure multidisciplinary scientific output.

7.1.1 Describing a GIScience Collaboration Network

To build a co-authorship network that can be taken as a window on the structure of GIScience by considering co-authorship as a surrogate measure of research collaborations.

For a long time, information scientists have being studying citations (Garfield, 1953; Price, 1965) and co-citation (Small, 1973) networks built from scientific literature published in academic journals, aimed at exploring, mapping and understanding the intellectual structure of sub-fields, fields and science in general (Boyack *et al*, 2005). However, as citations or co-citations are based on paper-paper relationships, their respective networks do not map actual contact between scientists, as paper 1 cites paper 2, does not imply that the authors knows each other.

In contrast, co-authorship measures imply to a certain extent a scientific acquaintance, as most people who have written a paper together will know one another quite well (Newman, 2001a). The results of exploring co-authorship networks in this study have revealed scientific collaborations in GIScience that are not the result of random interactions between scholars. It implies that not all scholars within the research community have the same probability of collaboration. The geographical analysis (in chapter 6) suggests that in GIScience, variables such as geographical proximity, language and cultural similarities increase the likelihood of establishing scientific collaborations. Moreover, results also show the *rich-get richer effect* (Barabasi and Bonabeau, 2003), as scholars in very few countries such as England, USA or Canada not only strengthened their GIScience collaboration partnership, but showed an increment in their international scientific links with colleagues from countries with less strong network participation patterns. As a result of the *rich-get richer effect,* the study identified a small group of individuals with very distinct and influential patterns of collaboration and co-authorship. Due to limitations of the bibliographic data authors could not be individually geo-referenced (see chapter 3), so it was not possible to establish the geographical location of this important group of scholars. The results have also revealed that the impact of the scientific contributions from this group of scholars on the discipline network topology varies according to the publication outlets. Sensitivity analysis (see Chapter 4) showed that their impact is greater in a co-authorship network covering scientific collaborations from strongly related topics to GIScience than in a network that covers more general topics.

To assemble the publication outlets that take into account the distinct characteristics of research in GIScience.

Subject categories in bibliographic databases cannot be taken as publication outlets for GIScience, as one category does not include all journals relevant to GIScience, as relevant publications to the discipline cannot be placed in one field or area. But also, one category may include other journals with less relevance to GIScience than others. The publication outlets representing GIScience are the result of identifying a core journal, and selecting other peripheral journals by quantifying a core-periphery distance in relation to the amount of work that cross-disciplinary authors have published. The more well-established authors have published in a peripheral journal, the closer the given journal is to the core. As a result, GIScience was represented using six different collaboration networks, built from publication outlets with different degrees of closeness between the core and the periphery. The proposed method avoided placing boundaries around GIScience and any allied discipline, or areas where GITechnologies can be applied.

7.1.2 Collaboration Networks

To evaluate the nature of the GIScience research collaboration network measuring and analyzing basic properties such as co-authorship, collaboration and authors' participation distributions, and topological features such as centrality measures and network's distances and connectivity.

GIScience has been presented as a multidisciplinary, emerging (UCGIS, 1996) interdisciplinary (Cova, 2000) or even a cross-disciplinary field (Mark, 1999) within which it is not possible to separate fundamental from applied research. Results from this study show that the closer the periphery is to the core, the closer is the network structure in representing scientific collaborations in GIScience. Two series of co-authorship networks (5-X and 4-X) were built. 5-X network series (5-2, 5-3 and 5-4) comprised core and peripheral journals with al least 2, 3 and 4 publications of well established authors (authors with at least 5 publications) in the core. Similarly, 4-X network series (4-2, 4-3, 4-4) comprised core and peripheral journals with al least 2,3 and 4 publications but by authors with more than 4 publication in the core. The notion of distance between core-periphery is inversely proportional to commonality of the sources among the well-established authors in the core. The commonest journals to many well-know authors the closer the journal will be to the core. Thus, distances between the core-periphery are shorter in 5-X than 4-X network.

The basic properties and statistics revealed similar low patterns of co-authorship and collaboration among the 5-X and 4-X networks. If work on GIScience is mainly the result of team work efforts (UGIS, 1996), the teams are not larger than 2 authors. However, on average, authors in 5-X were found to have less than 4.5 collaborators and around 5.0 in 4-X. Centrality measures based on the number of links between authors (degree, closeness and betweenness) identified that authors central in the 5-X series are less central in the 4-X networks. The list of central authors in the 4-X networks comprised unknown and known authors within the GIScience research community. In contrast, the 5-X network list showed a group of well-established authors in the GIScience community and from allied disciplines such as remote sensing and geography. However, the 4-X network structure is less centralized and its topology is less influenced by patterns of collaboration and co-authorship of central authors. In contrast, the 5-X series exhibited a more centralized network structure with a topology shaped by "*hubs*" authors.

This thesis showed that the closer the periphery is to the core, the better is the network structure in representing scientific collaboration in GIScience. The 5-X network of publication outlets represent GIScience collaboration as a network comprising core journals such as the *International Journal of Geographic Information Science*, *Geocomputation, Computer & Geosciences* and *Lecture Notes in Computer Science* that are linked to journals from very closed allied disciplines such as the *International Journal of Remote Sensing, Environment and Planning A* or *Progress in Human Geography* by a small group of scholars within the GIScience research community. In contrast, the 4-X networks were found to represent a very loose GIScience research community with larger distances between the periphery and the core resulting in the inclusion of less relevant journals in the network.

7.1.3 The Geography of the GIScience Core Collaboration Network

Is geographical closeness important for a GIScience researcher at the time to collaborate?

The spatial characterization of the GIScience scientific collaborations (represented by the 5-4 network) showed than 51% of the international collaborations included authors from 3 countries USA, England and Canada. It also revealed the emergence of countries such as China, Japan and Brazil as important new partners for GIScience scholars. However, the spatial GIScience collaboration network was found to be marked by exclusive symmetric partnerships between scholars from English speaking countries, cases such as Canada-England and Australia-USA, despite geographical distances. However this observation was severely limited by the data source being biased towards work in English, as documents in other languages were not included, giving, as Garcia-Ramon (2003) argues, privilege to the Anglophone scientific world.

The network was also found to exhibit asymmetrical relationships (Glänzel and Schubert, 2004) between authors from non-English speaker countries such as Mexico, Colombia, Costa Rica, Chile or Colombia with scholars in the USA and England that in many cases are much stronger than with neighbouring countries with which they even share the same language. In contrast, Scandinavian countries were found to hold an exclusive and symmetric relationship between themselves. Here cultural similarities and geographical proximity influence scientific collaborations, similar patterns found by

(Glänzel and Schubert, 2004) for global science networks. Moreover, results show that despite central European countries holding asymmetrical relationships among them, countries such as Germany and Belgium have strengthened their relationship with the USA, similar to the pattern found by Lian *et al* (2006) when studying scientific preference among 15 countries within the European Union.

7.1.4 Multidisciplinarity and GIScience

The results showed that individual patterns of collaboration and co-authorship such as the large number of book reviews (sole works) in *Progress in Human Geography, Environment and Planning A and Environment and Planning B – Planning and Design*, the high average of collaborators and team sizes in the *International Journal of Remote Sensing*, and the large share of works on remote sensing all helped to shape and influenced the structure of the collaboration network as a whole.

This study proposed measuring authors' multidisciplinarity by quantifying the distribution of their participation within each of the journals collected .Scaled Entropy (SE) identified three groups of authors: those who showed a concentration of their work In one journal only; those who exhibited a certain degree of scientific output distribution and those who exhibit high level of collaboration with a high number of the participant authors. The majority of authors that showed concentration (excluding authors with one paper) were found to be mostly contributors to remote sensing journals followed by those publishing in environmental and core GIScience issues. The largest percentage of authors who distributed their work was also found to be from remote sensing and environmental journals.

The results revealed a collaboration network with strong predominance patterns of collaboration and co-authorship of remote sensing authors (see Figure 7-54) who participated in large co-author teams but publishing mainly on remote sensing topics *(International Journal of Remote Sensing and Photogrammetric Engineering and Remote Sensing)*.

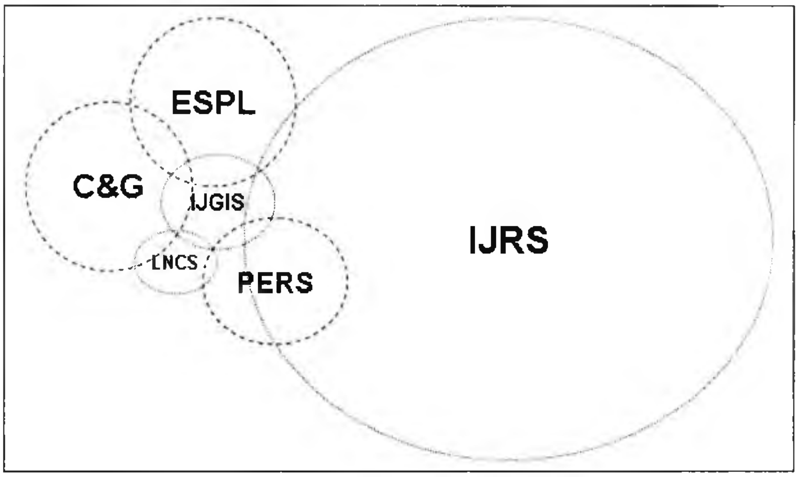


Figure 7-54 The share of each journal's co-authorship links in the GIScience largest interconnected sub-community.

More importantly is the fact that the collaboration network is also shaped and influenced by a small group of authors working on fundamental and applied GIScience issues (*International Journal of Geographic Information Science, Computers & Geosciences and Lecture Notes in Computing Science*). Contrary to remote sensing authors, "core GIScience" exhibit cross-multidisciplinary research interests and publication forums and a set of co-authorship connections that link smaller research communities to the GIScience core. Consequently, the studied co-authorship network represents GIScience scientific collaborations as a network with a core and a large group of smaller isolated sub-communities The core is shaped by a small group of GIScience scholars with a multidisciplinary set of collaborators linked to a highly clustered remote sensing research community. Outside the core, the network exhibits a large number of smaller sub-communities formed by authors with no co-authorship links to the core.

7.2 Evaluating the Proposed Approach

The proposed approach offered an alternative method for describing GIScience without using fixed subject categories defined by bibliographic databases such as ISI-WoK. The approach is limited to journals indexed by ISI- WoK. Results (section 5.1.2 chapter 5) revealed a significant number of book reviews especially among core and geoscience journals. It implies that academic books are important within the GIScience research community. Consequently, if scientific collaborations from GIScience books and other relevant journals (Transaction in GIScience or URISA) were included, the network's landscape would be shaped differently, less number of isolated authors and sub-communities. As more GIScience- related authors were part of the core, thus, it would be less dominated by remote sensing authors. Despite that, the current network structure manages to capture the nature of collaboration in GIScience revealing a centralized co-authorship structure around a group of authors identified as prominent scholars within the GIScience research community (see centrality and sensitivity analysis results, section 4.4.2 chapter 4).

In the case of the georeferencing process, limitations of the way that addresses are indexed also hindered the results. Only half of the papers in the publication outlet were collaborative works, thus 50% of the data source needn't be geo-referenced. Additionally, the large number of individual papers helped within the georeferencing process as there is not questioning that the address corresponds to the only co-authors. Overall, the georeferencing method was able to geo-reference in unambiguous manner around 60% of papers in the publication outlet. The other 40% are less certain cases where documents have more addresses than authors and more authors than address. The uncertainty introduced by both cases has not a large effect in the results of the georeferencing process as in the former case, not included co-authorship links would represent no longer valid collaborations(as the first author is no longer at that addresses). In the latter case, not included links due to the missing addresses would only lessen the strength of the collaboration.

7.3 Further Work

One characteristic showed by GIScience co-authorship network is low levels of authors' participation (large number of authors with one paper only). Researchers at the beginning of their carers may account for this, however, the study time window allowed the possibility of observing authors who had published a larger number of papers. Despite the fact that the average scientific output tends to be low (Lotka, 1926), the lack of inclusion by ISI-WoK of academic books, conference proceedings (GIScience conference proceedings are now indexed) and other important journals may have affected the network results. A solution could be to incorporate these missing sources and other important types of scientific collaboration such as workshops and conferences attendance, acknowledgements and research project bids. Additionally, due to the wide applicability of GISystems, the representation of collaboration in GIScience would benefit from the inclusion of relevant commercial projects.

The fact that not all works especially in peripheral journals are related to GIScience can also be the cause of low author participation and therefore authors who have not worked on GIScience may appear disconnected. One simple solution would be to compare the citation and co-citation indexes as complementary strategies for a highly dynamic selection of journals (Leydessdorff 2006a; Boyack *et al*, 2005). Thus, in that way one could establish a relationship between all journals and to evaluate the strength of their relations using co-factor analysis or multidimensional Scaling. Therefore, weaker relationships between journals that may be the cause of low author participations and a loose connected network core (that do not represent the disciplines adequately) can be excluded. Hence, only the journals that not only publish the work of top GIScience authors, but those who at the same time receive the approval of the community, measured by citation counts, could be included. In that way, one can be more sure that one-off relationships between authors outside the discipline are not taken into account.

This study is a first attempt to reveal geographical patterns of collaboration among scholars working on GIScience topics. However, the study can be extended to a finer level analysing patterns of collaboration at organizational level. As shown by Batty (2003) results within the same country patterns of citation varies from one research centre to another. In the same way in co-authorship, one might not expect to find the same patterns of scientific productivity between scholars working at different institutions within any one national setting. The spatial analysis at an organization level may well reveal further structure in GIScience collaboration.

Bibliography

Ahuja, M., and Carley, K., (1999). Network structure in virtual organizations. *Organization Science,* 10 (6), 741-57.

Albert , M., (2002). The Relevance of Pierre Bourdieu's Social Theory for the Study of Scientific Knowledge Production. *Canadian Journal of Sociology* CJS - Sociologists for a New Millennium [online] September - October 2002. Available from: http://www.cjsonline.ca/newmill/albert.html [Accessed 10 January 2007].

Albert, R., Jeong, H., and Barabasi, A., (1999). Diameter of the world-wide web. *Nature* , 401, 130-131.

Albert, R., and Barabasi, A., (2002). Statistical Mechanics of Complex Networks. *Reviews of Modern Physics*, 74, 47-97.

Abello, J., Pardalos, P.M., and Resende, M.G.C., (1999). On maximum clique problems in very large graphs. J. Abello and J. Vitter eds. In *External Memory Algorithms, AMS-DIMACS Series on Discrete Mathematics and Theoretical Computer Science*, 50, 119-130.

Adamic. L., and Adar, E. (2005). How to search a social network. Preprint submitted to *Social Networks*. 27(3), 187-20.

Adamic, L.A., and. Huberman, B.A., (2002) Zipf's law and the Internet', *Glottometrics,* 3, 143-150.

Amaral, L. A. N., Scala, A., Barthelemy, M., and Stanley, H.E., (2000). Classes of small-world networks. *Proceedings of the National Academy of Sciences U.S.A (PNAS)*, 97(21), 11149-11152.

Arciniegas, C., and Wood, J., (2006). Collaboration Networks Revealed by IJGIS Authors. In : P. Fisher, ed. *Classics from IJGIS: Twenty years of the International Journal of Geographical Information Science and Systems.* London : Taylor & Francis, 547-553.

Archambault, E., (2000). Comments to the article by Rousseau & Rousseau, *Cybermetrics* [online], 4 (1), correspondence. Available from: http://www.cindoc.csic.es/cybermetrics/articles/v4i1c1.html [Accessed 19 July 2006].

Barabasi, A., (2002). Linked - The new science of networks. Cambridge, USA: Perseus Publishing, 1-256.

Barabasi, A., and Bonabeau, E., (2003). Scale-Free Networks. *Scientific American*, 50-59.

Barabasi, A., and Albert, R., (1999). Emergence of scaling in random networks. *Science*, 286, 509-512.

Barabasi, A., Jeong, H., Néda Z., Ravasz E., Schubert A. and Vicsek, T., (2002). Evolution of the social network of scientific collaborations. *Physica A*, 311, 590-614.

Batty, M., (2003). Citation Geography: It's About Location. *The Scientist,* 17(16), 10-12.

Batty, M., (2003a). The Geography of Scientific Citation. *Environmental and Planning A*, 35, 761-770.

Bensman, S.J., (2001). Bradford's Law and Fuzzy Sets: Statistical Implications for library Analysis. *IFLA Journal*, 27, 238-246.

Bordons, M., and Gomez, I., (2000). Collaboration Networks in Science. In: Medford, N.J, ed, A *festschrift in Honour of Eugene Garfield. Information* Medford, NJ : Today, 197-213.

Borgatti, S.P., (2004). Introduction to Graph Theory. In Social Network Analysis Instructional Web Site. Available from http://www.analytictech.com/networks/graphtheory.htm. [02 Feb 2004].

Borgatti, S.P., (2005). Centrality and Network Flow. *Social Networks*, 25 (1), 55-71.

Borgman C.L and Furner, J., (2002). Scholarly Communication and Bibliometrics. *In*: B, Cronin, ,ed. *Annual Review of Information Science* and *Technology: vol. 36*. Medford, N.J.: Information Today, Inc, 3-72.

Boyack, K., Klavans, R. and Börner, K., (2005). Mapping the backbone of science. *Scientometrics*, 64(3), 351-374.

Bozeman, B. and Lee, S., (2005). The Impact of Research Collaboration on Scientific Productivity. *Social Studies of Science*, 35(5), 673-702.

Börner, K., Maru, J., and Goldstone, R., (2004). The Simultaneous Evolution of Author and Paper Networks. *Proceedings of the National Academy of Sciences U.S.A (PNAS)*. 101(1), 5266-5273.

Börner, K., Dall'Asta, L., Ke, W. and Vespignani, A. (2005). Studying the Emerging Global Brain: Analyszing and Visualizaing the Impact of Co-Authorship Teams. *Complexity*. 10(4), 57-67.

Broadus, R.M., (1987). Toward a definition of bibliometrics. *Scientometrics,* 12(5 -6), 373-379.

Bradford, S.C., (1934). Sources of Information on Specific Subjects. *Engineering: An Illustrated Weekly Journal* , 137, 85-86.

Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., Tomkins, A., and Wiener, J., (2000). Graph structure in the web, *Computer Networks,* 33, 309-320.

Buchanan, M., (2002). Small world: uncovering nature's hidden networks. The Orion Publishing Group: London , 1-222.

Bührer, S., (2002). Network Analyses. In G., FAHRENKROG, W., POLT, J., ROJO, A. , TÜBKE, and K., ZINÖCKER eds. *RTD-Evaluation Toolbox Assessing the Socio-Economic Impact of RTD-Policies*. European Commission - Joint Research Centre. Institute for Prospective Technological Studies (IPTS), Seville, Spain, 163-169.

Calero, C., Buteri, R., Cabello-Valdes, C., and Noyons, Ed., (2006). How to identify research groups using publications analysis: an example in the field of nanotechnology. *Scientometrics*, 66(2), 365-376.

Carrier, R., (2005). Entropy explained. Addendum to "Bad Science, Worse Philosophy: the Quackery and Logic-Chopping of David Foster's - The Philosophical Scientists". Available from: http://www.infidels.org/library/modern/richard_carrier/entropy.html [Accessed 10 June 2007].

Carvalho, R., and Batty, M., (2006). The Geography of Scientific Productivity: Scaling in U.S. Computer Science. *Journal of Statistical Mechanics: Theory and Experiment*. 2006(10), P10012.

Chapman, S. (2005). SinMetrics Open Source Library. Available from: http://www.dcs.shef.ac.uk/~sam/stringmetrics.html#dice [Accessed 10 February 2007].

Cohen, W.W., Ravikumar, P., Fienberg, S.E., (2003). A comparison of string distance metrics for name-matching tasks. In: *Proceedings of the IJCAI-2003 Workshop on Information Integration on the Web*, Acapulco, Mexico, 73-78.

Costas, R. and Bordons, M., (2005). Methodological procedure to overcome the lack of normalization of author names in bibliometrics analyses at micro level. In: *Proceedings of 10th International Conference of the International Society for Scientomet4rics and Informetrics. 24-28 July, Stockholm, Sweden*. Stockholm: Karolinska University Press, Karolinska Institute, 688-689.

Cornwell, B., (2005). A complement-derived centrality index for disconnected graphs. *Connections*, 26(2), 72-83.

Cova, T., (2000). *GIScience 2000: Report from Savannah or "Midnight in the Garden of Goodchild and Egenhofer"* [online]. Department of Geography, University of

Utah. Available from: http://www.giscafe.com/technical/special_events/cova_GIScience2000.php [Accessed 10 January 2007].

Crane, D., and Small, H., (1992). American sociology since the seventies: the emerging identity crisis in the discipline. In T.C. HALLIDAY and M. JANOWITZ eds. *Sociology and its Publics: The Forms and Fates of Disciplinary Organization*). Chicago: University of Chicago Press, 197–234.

Cronin, B., (2001). Hyper-authorship: a post-modern perversion or evidence of a structural shift in scholarly communication practices?, *Journal of the American Society for Information Science and Technology,* 52(7), 558-569.

De Castron, R., and Grossman, J.W., (1990). Famous trials to Paul Erdős. *Mathematical Intelligencer*, 21, 51-63.

De Nooy, W., Mrvar, A., and Batagelj, V., (2005). Exploratory Social Network Analysis with Pajek. In Structural Analysis in the Social Sciences. Cambridge , NY: Cambridge University Press, 123-137.

Dogan, M., (2001). Specialization and Recombination of Specialties in the social sciences. IN: N. J. SMELSER and P.B. BALTES, eds. *International Encyclopaedia of Social and Behavioural Sciences.*. London: Pergamon-Elsevier Science, 14851-14855.

Ding, Y., Foo, S., and Chowdhury, G., (1999). A bibliometric analysis of collaboration in the Field of Information Retrieval. *International Information & Librarian Review*, 30, 367-376.

Ebel, H., Davidsen J. and Bornholdt, S., (2003). Dynamics of Social Networks. Complexity –, 8(2), 24-27.

Elmacioglu, E., and Lee, D., (2005). On Six Degrees of Separation in DBLP-DB and More. *SIGMOD Record*, 34(2), 33-40.

Erdos, P., and Renyi, A., (1959). On Random Graphs. *Publicationes Mathematicae,* 6, 290-297.

Faloutsos, M., Faloutsos, P., and Faloutsos, C., (1999). On Power-Law Relationships of the Internet Topology. *Computing . Communication. Review.* 29, 251-262.

Fisher, P.F., (1998). Geographic Information Systems and Geographic Information Science: A Case of the Wrong Metaphor. In: Proceedings of the 8th International Symposium on Spatial Data Handling. Simon Fraser University, Vancouver, pp. 321–330.

Fisher, P.F., (2001). Citations to the International Journal of Geographical Information Systems and Science: the first 10 years. *International Journal of Geographical Information Science*, 15(1), 1-6.

Fisher, P., (2006). Introduction – Twenty years of IJGIS: Choosing the Classics. In : P. Fisher, ed. *Classics from IJGIS: Twenty years of the International Journal of Geographical Information Science and Systems.* Taylor & Francis: London, 1-6.

Fisher, P.F., and Unwin, D., (2005). Re-presenting Geographical Information Systems. P. Fisher, ed. IN *Re-presenting GIS*. England: Wiley, 1-17.

Freeman, L.C., (1979). Centrality in networks: I. Conceptual clarification. *Social Networks,* 1, 215–239.**Frenken, K., and Leydesdorff, L.A.,** (2004). Scientometrics and the evaluation of European integration. IN: T. BROWN and J. ULIJN, eds. *Innovation, Entrepreneurship and Culture: The Interaction between Technology, Progress and Economic Growth*. Cheltenham, UK: Edward Elgar Publishing, 87-102.

Garcia-Ramon, M. (2003). Globalization and international geography: the question of languages and scholarly traditions. *Progress in Human Geography*, 27(1), 1-5.

Garfield, E., (1955). Citations Indexes for science. *Science, New Series*, 122 (3159), 108-111.

Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P., and Trow, M., (1994). The new production of Knowledge. London: Sage, 17-45.

Glänzel, W., (2003). Bibliometrics As A Research Field: A course on theory and application of bibliometric indicators. Available from: http://www.norslis.net/2004/Bib_Module_KUL.pdf. [Accessed 1st December 2006].

Glänzel, W., and Schubert, A., (2004). Analyzing Scientific Networks Through Co-Authorship. *In*: H.F. Moed, U. Schmoch and W. Glänzel, eds. *Handbook of Quantitative Science and Technology Research: The Use of Publication and Patent Statistics in Studies of S&T Systems*. Dordrecht, The Netherlands: Kluwer Academic Publishers, 257-276.

Goodchild, M.F., (1991). Just the facts. *Political Geography Quarterly* , 10, 335–37.

Goodchild, M.F., (1992). Geographical Information Science. *International Journal of Geographical Information Systems*. 6(1), 31-45.

Goodchild, M.F., (2004). The Validity and Usefulness of Laws in Geographic Information Science and Geography. *The Annals of the Association of American Geograhpers*, 94(2), 300-303.

Goodchild, M.F., (2006). Geographical Information Science: Fifteen Years Later. In: P. Fisher, ed. *Classics from IJGIS: Twenty years of the International Journal of Geographical Information Science and Systems*. Taylor & Francis: London, 199-203.

Goodchild, M.F., Egenhofer, M.J., Kemp, K., Mark, D., and Sheppard, E., (1999). Introduction to the Varenius Project. *International Journal of Geographical Information Systems*. 13(8), 731-745.

Granovetter, M., (1973). The Strength of Weak Ties. *American Journal of Sociology*, 78(6), 1360-1380.

Griffith, B.C., (1988). Derek's Price Puzzles; Numerical Metaphors for the Operations of Science, *Science, Technology, & Human Values*, 13(3-4), 351-360.

Grossman, J.W., and Ion, P.D.F., (1995). On a portion of the well-known collaboration graph, *Congressus Numerantium*, 108, 129-131.

Guare, J., (1990). Six Degrees of Separation: A Play. Vintage, New York.

Gutierrez, P., and Lopez-Nieva, P., (2001). Are international journals of human geography really international?. *Progress in Human Geography*, 25(1), 53-69.

Hagen, G., Killinger, D.K., and Streeter, R.B., (1997). An Analysis of Communication Networks Among Tampa Bay Economic Development Organizations, *Connections*, 20(2), 13-22.

Hayes, B., (2000). Graph Theory in Practice: Part I. *American Scientist*. 88(1), 9-13.

Havemann, F., Heinz, M., and Kretschmer, H., (2006). Collaboration and distances between German immunological institutes – A trend analysis. *Journal of Biomedical Discovery and Collaboration*, 1, 1-6.

Horn, D.B., Finholt, T., Birnholtz, J., Motwani, D., and Jayaraman, S., (2004). Six Degrees of Jonathan Grudin: A social Network Analysis of the Evolution and Impact of CSCW Research. *Letters Chi – ACM*, 6(3), 582-591.

Katz, J.S., (1999). Bibliometric Indicators and the Social Sciences [online]. SPRU: Science and Technology Policy Research, University of Susex. Available from: http://www.sussex.ac.uk/Users/sylvank/pubs/ESRC.pdf [Accessed 20 January 2006].

Katz, J.S., and Martin, R.B., (1997). What is research collaboration? *Research Policy*, 26, 1-18.

Klavans, R., and Boyack, K.W. (2006). Identifying a Better Measure of Relatedness for Mapping Science. *Journal of the America Society for Information Science and Technology*. 57 (2), 251-263.

Kennedy, M., (1994). Review of Geographical Information Systems: Principles and Applications. *Annals of the Association of American Geographers*, 84, 172-173.

Kleinfeld, J.S., (2002). The small-world problem. *Society*, 39(2), 61-66.

Kocken, M., (1987). How well do we acknowledge intellectual debts? *Journal of Documentation*, 43 (1), 54-64.

Koku, E., and Wellman, B., (2002). Scholarly Network as Learning Communities: The Case of TechNet. IN S Barab and R Klingl, eds. *Designing Virtual Communities in the Service of Learning.* Cambridge, NY: Cambridge University Press, 299-338.

Krebs, V.E., (2002). Uncloaking Terrorist Networks, *First Monday*, 7(4). Available from: http://www.firstmonday.dk/issues/issue7_4/krebs/index.html [Accessed 5 January 2007].

Laudel, G., (2002). What do we measure by co-authorships?, *Research Evaluation*, 11(1), 3-15.

Leydesdorff, L., (2006). Betweenness Centrality as an Indictor of the Interdisciplinary of Scientific Journals. *Conference on Science and Technology Indicators*, 7-9 September 2006 Leuven, Belgium.

Leydesdorff, L., (2006a). Mapping Interdisciplinary at the Interfaces between the Science Citation Index and the Social Science Citation Index. *Amsterdam School of Communication Research, University of Amsterdam.* Available from: http://users.fmg.uva.nl/lleydesdorff/sci_sosci/index.htm [Accessed 7 November 2006].

Leydessdorff, L., (2006b). Can Scientific Journals be Classified in terms of Aggregated Journal-Journal Citations Relations using the Journal Citations Reports?. *Journal of the America Society for Information Science and Technology*. 57(5), 601-613.

Leydessdorff, L., and Cozzens, S., (1993). The delineation of Specialities in Terms of Journals Using the Dynamic Journal Set of the SCI. *Scientometrics*, 26(1), 135-156.

Liu, X., Bollen, J., Nelson, M. L., and Van de Sompel, H., (2005). Co-Authorship Networks in the Digital Library Research Community. *Information Processing & Management*, 41(6), 1462-1480.

Liang, L., Zhang, L., Kretschmer, H., Scharnhorst, A. (2006). Geographical and Lingual Preferences in Scientific Collaboration of the European Union (1994-2003). In: Proceedings International Workshop on Webometrics, Informetrics and Scientometrics and Seventh COLLNET Meeting, 10-12 May, 2006, Nancy , France.

Lokta, A.J., (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences*, 16(12), 317-323.

Logan, E.L., Tallahassee, F.L and Pao M. L.,. (1991). Identification of Key Authors in a Collaborative Network. *Proceedings of the Asis Annual Meeting* 28, 261-266.

Longley, P., Goodchild, M., Maguire, D., and Rhind, D., (2001). *Geographic Information Systems and Science*. England: Wiley, 1st Edition, 1-80.

Mark, D.M., (1999). Geographic Information Science: Critical Issues in an Emerging Cross-Disciplinary Research Domain. Workshop Report. Available from: http://www.ncgia.buffalo.edu/GIScienceReport.html [Accessed 01 November 2006].

Mark, D.M., (2003). Geographic Information Science: defining the field. In, I.M. Duckhan, M.F. Goodchild , and M.F. Worboys, eds. *Foundations of Geographic Information Science*, New York: Taylor & Francis, 3-18.

Melin, G., and Persson, O., (1996). Studying research collaboration using co-authorships. *Scientometrics*, 36(3), 363-377.

Mikhailov, A. I., (1984). Scientific communication. In A.I. Mikhailov, A.I. Chernyi and R.S. Giliarevskii., eds. *Scientific communication and informatics*. Information Arllington, VA: Resources Press.

Milgram, S., (1967). The small world problem. *Psychology Today*, 2, 60-67.

Mizruchi, M.S., (1994). Social Network Analysis: Recent Achievements and Current Controversies. *Acta Sociologica.* 37, 329-343.

Moody, J., (2004). The Structure of a Social Science Collaboration Network: Disciplinary Cohesion from 1963 to 1999. *American Sociological Review*, 69, 213–238.

Morillo, F., Bordons, M., and Gomez, I., (2003). Interdisciplinary in Science: A Tentative Typo logy of Disciplines and Research Areas. *Journal of the America Society for Information Science and Technology*, 54(13), 1237-1249.

Murray, C., Ke, W., and Börner, K., (2006). Mapping Scientific Disciplines and Author Expertise Based on Personal Bibliography Files. *IEEE*, 258-263.

Narin, F., Carpenter, M.P., and Woolf, P., (1984). Technological Performance Assessments Based on Patents and Patent Citations. *IEEE Transactions on Engineering Management,* 31 (4), 172-183.

Nascimiento, M.A., Sander J.. and Pound, J., (2003). Analysis of SIGMOD's Co-Authorship Graph. *ACM SIGMOD*, 32(3), 8–10.

Newman, M.E.J., (2000). Comments to the article by Rousseau & Rousseau *Cybermetrics* [online], 4 (1), correspondence. Available from: http://www.cindoc.csic.es/cybermetrics/articles/v4i1c2.html [Accessed 19 July 2006].

Newman, M.E.J., (2001a). The structure of scientific collaborations networks . *Proceedings of the National Academy of Sciences U.S.A (PNAS)*, 98, 404-409.

Newman, M.E.J., (2001b). Scientific collaboration networks. I. Network construction and fundamental results. *Physical Review E*, 64, 016131-8.

Newman, M.E.J., (2001c). Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Physical Review*, 64, 016132-7.

Newman, M.E,J., (2003). The Structure and Function of Complex Networks. SIREV. 45(2), 167 – 256.

Newman, M.E.J., (2004b). Who is the best connected Scientist? A study of scientific co-authorship networks. *Lecture Notes in Physics*, vol. 650, p.337-370 (2004).

Newman, M.E.J., (2004c). Co-authorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences U.S.A (PNAS).* 101(1), 5200-5205.

Newman, M.E.J., (2005) Power laws, Pareto distribution and Zip's law. *Contemporary Physics*. 46(5), 323-351.

Newman, M.E.J., (2005a). A measure of betweenness centrality based on random walks *Social Networks*, 27 (1), 39-54.

Newman, M.E.J, Barabasi, A. And Watts, D.J., (2006). Chapter One – Introduction. IN M.E.J., Newman, A Barabasi and DJ Watts, eds. *The structure and Dynamics of Networks*. Princeton, NJ: Princeton University Press, 1-26.

Newman, M.E,J., Watts D.J. and Strogatz, S.H. (2002). Random graph models of social networks. *Proceedings of the National Academy of Sciences U.S.A (PNAS).* 99 (1), 2566-2572.

Nowotny, H., Scott, P., And Gibbons, M., (2003). Mode 2' Revisited: The New Production of Knowledge. *Minerva,* 41(3), 179–194.

Openshaw, S., (1991). A view on the GIS crisis in geography, or, using GIS to put Humpty-Dumpty back together again. *Environment and Planning A*, 23, 621–28.

Otte, E., and Rousseau, R., (2002). Social network analysis: a powerful strategy, also for the information sciences. *Journal of Information Science*, 28(6), 441-453.

Pao, M.L., (1986). An empirical examination of Lotka's Law. *Journal of the American Society for Information Science*, 26-33.

Pierce, S. J., (1991). Subject areas, disciplines and the concept of authority. *LISR - Library and Information Science Research*, 13, 21-35.

Peterson, I., (1998). Close Connections- It's a small world of crickets, nerve cells, computers, and people. *Science News*. 154(8), 124-128.

Pool, I de Sola., and Kochen, M., (1978). Contacts and Influence, *Social Networks*, 1(1), 5-51.

Price, D. J. de S., (1965). Network of Scientific Papers. *Science* 149, 510-515.

Pritchard, A., (1969). Statistical bibliography or Bibliometrics?, *Journal of Documentation,* 24, 348-349.

Ramamurthi, K., (2004). Königsberg Bridges Problem. Available from: http://math.ucsd.edu/~rramamu/app/CSUSMtalk3.html. [Accessed 13 Jan 2004].

Rey-Rocha, J., and Martin-Sempere, M.J., (2004). Patterns of the foreign contributions in some domestic vs. international journals on Earth Science. *Scientometrics*, 59(1), 95-115.

Rousseau , B., and Rousseau, R., (2000). LOTKA: A program to fit a power law distribution to observed frequency data. *Cybermetrics* [online], 4 (1), paper 4. Available from: http://www.cindoc.csic.es/cybermetrics/articles/v4i1p4.html [Accessed 19 July 2006].

Shannon, C.E., (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 379-423 and 623-656, July and October.

Small, H.G., (1973). Co-citation in the scientific literature; a new measure of the relationship between two documents. *Journal of the American Society for Information Science*, 24, 265–269.

Stichweh, R., (2001). Scientific Disciplines, History of. IN: N. J. SMELSER and P.B. BALTES, eds. *International Encyclopaedia of Social and Behavioural Sciences*, 13727-13731.

Stephenson, K., and Zelen, M., (1989). Rethinking centrality: Methods and examples, *Social Networks*, 1(11), 1-37.

Subramanyam, K., (1983). Bibliometric studies of research collaboration: A review. *Journal of Information Science* 6, 33-38.

Schuurman, S., (2000). Trouble in the heartland: GIS and its critics in the 1990s. *Progress in Human Geography*, 24 (4), 569–590.

Taylor, P.J., (1990). GKS- Geographic Knowledge Systems. *Political Geography Quarterly*, 9, 211-12.

Travers, J., and Milgram, S., (1969). An experiment study of the small world problem. *Sociometry*, 32(4), 425-443.

Tobler, W., (1970). A Computer Movie Simulating Urban Growth in the Detroit Region, *Economic Geography,* 46 (2), 234-240).

Tobler, W., (2004). On the First Law of Geography: A Reply. *Annals of the Association of American Geographers*, 94(2), 304-310.

University Consortium for Geographic Information Science, (UCGIS)., (1996). Research Priorities for Geographic Information Science. *Cartography and Geographic Information Systems*. 23(3), 1-15.

Van den Besselaar, P., and Heimeriks, G., (2001). Disciplinary, Multidisciplinary, Interdisciplinary – Concepts and Indicators. In M. DAVIS & C.S. WILSON, eds. *Proceedings of the 8th International Conference on Scientometrics and Informetrics, 16-20 July, 2001 Sydney, Australia.* Sydney: Bibliometric and Informetric Research Group (BIRG), University of New South Wales (UNSW), 705-716.

Von Unbegrn-Sternberg, S., (2000). Scientific communication and bibliometrics [online]. Abo Akademi Univesity, Department of Information Studies. Available from: http://www.abo.fi/~sungern/comm00.htm [Accessed 10 March 2005].

Wagner, C., and Leydesdorff, L., (2003). Mapping the network of global science: comparing international co-authorships from 1990 to 2000. International Journal of Technology and Globalisation. 1(2). 185-208.

Watts, D. J., and Strogatz, S. H., (1998). Collective dynamics of `small-world' networks, *Nature*, 393, 440-442.

Watts, D.J., (2003). Six Degrees - The science of a connected age. Vintage: London, 1-368.

Wheterell, C., Plankas, A., and Wellman, B., (1994). Social networks, kinship, and community in Eastern Europe. *Journal of Interdisciplinary History*, 24, 639-663.

Whitley, R., (2000). The Intellectual and Social Organization of the Sciences. 2nd ed. Oxford: Oxford University Press, 153-265.

Wood, J.D., (1996). The geomorphological characterisation of digital elevation models. Thesis (PhD). University of Leicester, UK.

Wright, D., Goodchild, M., and Proctor, J., (1997). Demystifying the Persistent Ambiguity of GIS as "Tool" Versus "Science". *The Annals of the Association of American Geographers*, 8(2), 346-362.

Zipf, G.K., (1930). Relative frequency as a determinant of phonetic change. *Language*, 6 (1), 86-88. Appendices.

Appendices

Appendix A.Full list of journals where well-established authors in IJGIS have published at least one paper.

Journal Name	Total No. Authors	Journal Name	Total No. Authors
Environment And Planning B-Planning & Design	18	Climate Research	1
Computers & Geosciences	13	Communications In Algebra	1
Environment And Planning A	13	Communications In Statistics-Simulation And Computation	1
International Journal Of Remote Sensing	10	Computers & Education	1
Photogrammetric Engineering And Remote Sensing	10	Computers & Graphics-Uk	1
Progress In Human Geography	9	Conservation Biology	1
Annals Of The Association Of American Geographers	8	Decision Support Systems	1
Computers Environment And Urban Systems	8	Discrete Dynamics In Nature And Society	1
Geographical Analysis	7	Discrete Geometry For Computer Imagery Lecture Notes In Computer Science	1
Applied Geography	6	Ecology	1
Cartographic Journal	6	Economic Geography	1
Geoinformatica	6	Environment And Behavior	1
Journal Of Geography In Higher Education	6	Environment And Planning D-Society & Space	1
Spatial Information Theory Lecture Notes In Computer Science	6	Environmental & Resource Economics	1
Advances In Spatial Databases Lecture Notes In Computer Science	5	Environmental Modelling & Software	1
Area	5	Environmental Monitoring And Assessment	1
Geographical Journal	5	Epidemiology	1
Geography	5	European Journal Of Public Health	1
Journal Of The Royal Statistical Society Series D-The Statistician	5	European Urban And Regional Studies	1
Transactions Of The Institute Of British Geographers	5	Geoforum	1
Cartography And Geographic Information Systems	4	Geophysical Research Letters	1
Computer Journal	4	Global Environmental Change-Human And Policy Dimensions	1
Earth Surface Processes And Landforms	4	Graphical Models	1
Environment And Planning C-Government And Policy	4	Health & Place	1
Fuzzy Sets And Systems	4	Housing Studies	1
Regional Studies	4	Hydrobiologia	1
Urban Studies	4	Ieee Computer Graphics And Applications	1
Computers & Graphics	3	Ieee Transactions On Geoscience And Remote Sensing	1
Environmental And Ecological Statistics	3	Ieee Transactions On Neural Networks	1

Geoderma	3	Ieee Transactions On Pattern Analysis And Machine Intelligence	1
Geomorphology	3	Ieee Transactions On Visualization And Computer Graphics	1
Journal Of Public Health Medicine	3	Information Sciences	1
Journal Of Regional Science	3	Information Technology And Libraries	1
Journal Of The American Planning Association	3	Integrated Spatial Databases: Digital Images And Gis Lecture Notes In Computer Science	1
Journal Of Visual Languages And Computing	3	International Journal Of Computer Mathematics	1
Professional Geographer	3	International Journal Of Human-Computer Studies	1
Spatial Information Theory: A Theoretical Basic For Gis. Lecture Notes In Computer Science	3	International Regional Science Review	1
Technometrics	3	Journal Of Agricultural Economics	1
Visual Computer	3	Journal Of Applied Meteorology	1
Acm Transactions On Graphics	2	Journal Of Archaeological Science	1
Advances In Spatial And Temporal Databases, Proceedings. Lecture Notes In Computer Science.	2	Journal Of Construction Engineering And Management-Asce	1
Agriculture Ecosystems & Environment	2	Journal Of Glaciology	1
Canadian Geographer-Geographe Canadien	2	Journal Of Retailing	1
Catena	2	Journal Of Soil And Water Conservation	1
Cities	2	Journal Of Soil Science	1
Computational Statistics	2	Journal Of Statistical Computation And Simulation	1
Computer-Aided Design	2	Journal Of Statistical Planning And Inference	1
Cvgip-Graphical Models And Image Processing	2	Journal Of Theoretical Biology	1
Ecological Modelling	2	Journal Of Urban Economics	1
Environmental Management	2	Journal Of Vegetation Science	1
Forestry	2	Journal Of Wildlife Management	1
Futures	2	Kidney International	1
Habitat International	2	Landscape Ecology	1
Hydrological Processes	2	Lecture Notes In Artificial Intelligence	1
Ieee Multimedia	2	Lecture Notes In Computer Science	1
Ieee Transactions On Knowledge And Data Engineering	2	Lecture Notes In Economics And Mathematical Systems	1
International Journal Of Climatology	2	Linear Algebra And Its Applications	1
Interoperating Geographic Information Systems Lecture Notes In Computer Science	2	Mathematical Social Sciences	1
Isprs Journal Of Photogrammetry And Remote Sensing	2	Mitteilungen Der Osterreichischen Geographischen Gesellschaft	1
Journal Of Environmental Management	2	Nature	1
Journal Of Epidemiology And Community Health	2	Nutrient Cycling In Agroecosystems	1
Journal Of Hazardous Materials	2	Papers In Regional Science	1
Journal Of Hydrology	2	Physical Review E	1
Journal Of Information Science	2	Plant And Soil	1
Land Use Policy	2	Political Geography Quarterly	1

Landscape And Urban Planning	2	Proceedings Of The Edinburgh Mathematical Society	1
Papers In Regional Science	2	Progress In Physical Geography	1
Pattern Recognition Letters	2	Progress In Planning	1
Regional Science And Urban Economics	2	Public Health	1
Remote Sensing Of Environment	2	Quarterly Journal Of Engineering Geology	1
Risk Analysis	2	Quaternary Science Reviews	1
Transportation Research Part C-Emerging Technologies	2	Science & Technology Libraries	1
Acm Transactions On Information Systems	1	Scientist	1
Advances In Case-Based Reasoning Lecture Notes In Artificial Intelligence	1	Scottish Geographical Magazine	1
African Journal Of Ecology	1	Sedimentary Geology	1
Algorithmica	1	Social Science & Medicine	1
American Naturalist	1	Soil & Tillage Research	1
Annals Of Mathematics And Artificial Intelligence	1	Soil Science Society Of America Journal	1
Annals Of Operations Research	1	Spatial Cognition Ii Lecture Notes In Computer Science	1
Annals Of Regional Science	1	Spatio-Temporal Databases: The Chrochronos Approach Lecture Notes In Computer Sciene	1
Architectural Design	1	Statistics & Probability Letters	1
Aslib Proceedings	1	Survey Review	1
Biodiversity And Conservation	1	Topology And Its Applications	1
British Medical Journal	1	Transport Reviews	1
Canadian Journal Of Soil Science	1	Urban Geography	1
Chemosphere	1	Water Air And Soil Pollution	1
		Water Resources Research	1

Appendix B. Co-authorship Networks and their data sources.

	JOURNAL TITLE	(4, 2)	(4,3)	(4,4)	(5, 2)	(5,3)	(5,4)
1	Acm Transactions On Graphics						
2	Agriculture Ecosystems & Environment						
3	Annals Of The Association Of American Geographers						
4	Applied Geography						
5	Area						
6	Canadian Geographer-Geographe Canadien						
7	Cartographic Journal						
8	Cartography And Geographic Information Systems						
9	Catena						
10	Cities						
11	Computational Statistics						
12	Computer Journal						
13	Computer-Aided Design						
14	Computers & Geosciences						
15	Computers & Graphics						
16	Computers Environment And Urban Systems						
17	Cvgip-Graphical Models And Image Processing						
18	Earth Surface Processes And Landforms						
19	Ecological Modelling						
20	Environment And Planning A						
21	Environment And Planning B-Planning & Design						
22	Environment And Planning C-Government And Policy						
23	Environmental And Ecological Statistics						
24	Environmental Management						
25	Forestry						
26	Futures						
27	Fuzzy Sets And Systems						
28	Geoderma						
29	Geographical Analysis						
30	Geographical Journal						
31	Geography						
32	Geoinformatica + Transaction in GIS[1]						
33	Geomorphology						
34	Habitat International						
35	Hydrological Processes						
36	IEEE Multimedia						
37	IEEE Transactions On Knowledge And Data Engineering						
38	International Journal Of Climatology						
39	International Journal Of Geographical Information Science						
40	International Journal Of Remote Sensing						
41	ISPR Journal of Photogrammetry And Remote Sensing						
42	Journal Of Environmental Management						
43	Journal Of Epidemiology And Community Health						

44	Journal Of Geography In Higher Education						
45	Journal Of Hazardous Materials						
46	Journal Of Hydrology						
47	Journal Of Information Science						
48	Journal Of Public Health Medicine						
49	Journal Of Regional Science						
50	Journal Of The American Planning Association						
51	Journal Of The Royal Statistical Society Series D-The Statistician						
52	Journal Of Visual Languages And Computing						
53	Land Use Policy						
54	Landscape And Urban Planning						
55	Lecture Notes In Computer Sciences[2]						
56	Papers In Regional Science						
57	Pattern Recognition Letters						
58	Photogrammetric Engineering And Remote Sensing						
59	Professional Geographer						
60	Progress In Human Geography						
61	Regional Science And Urban Economics						
62	Regional Studies						
63	Remote Sensing Of Environment						
64	Risk Analysis						
65	Technometrics						
66	Transactions Of The Institute Of British Geographers						
67	Transportation Research Part C-Emerging Technologies						
68	Urban Studies						
69	Visual Computer						

Appendix C. List of journals and their contributing (in no. of papers) to 4-X co-authorship networks series

JOURNAL TITLE	4-4	4-3	4-2
Acm Transactions On Graphics			234
Agriculture Ecosystems & Environment			1297
Annals Of The Association Of American Geographers	1183	1183	1183
Applied Geography	463	463	463
Area	1094	1094	1094
Canadian Geographer-Geographe Canadien			660
Cartographic Journal	352	352	352
Cartography And Geographic Information Systems	56	56	56
Catena			672
Cities			635
Computational Statistics			186
Computer Journal	715	715	715
Computer-Aided Design			915
Computers & Geosciences	1223	1223	1223
Computers & Graphics		562	562
Computers Environment And Urban Systems	169	169	169
Cvgip-Graphical Models And Image Processing			135
Earth Surface Processes And Landforms	914	914	914
Ecological Modelling			1971
Environment And Planning A	2180	2180	2180
Environment And Planning B-Planning & Design	1019	1019	1019
Environment And Planning C-Government And Policy	583	583	583
Environmental And Ecological Statistics		186	186
Environmental Management			1075
Forestry			394
Futures			151
Fuzzy Sets And Systems	3393	3393	3393
Geoderma		1137	1137
Geographical Analysis	267	267	267
Geographical Journal	1198	1198	1198
Geography	1240	1240	1240
Geoinformatica	118	118	118
Geomorphology		970	970
Habitat International			83
Hydrological Processes			1435
IEEE Multimedia			377
IEEE Transactions On Knowledge And Data Engineering			867
International Journal Of Climatology			1014
International Journal Of Geographical Information Science	626	626	626
International Journal Of Remote Sensing	3015	3015	3015
ISPR Journal of Photogrammetry And Remote Sensing			363
Journal Of Environmental Management			764
Journal Of Epidemiology And Community Health			1725
Journal Of Geography In Higher Education	477	477	477
Journal Of Hazardous Materials			1360

Journal Of Hydrology			2655
Journal Of Information Science			571
Journal Of Public Health Medicine		1222	1222
Journal Of Regional Science		781	781
Journal Of The American Planning Association		1367	1367
Journal Of The Royal Statistical Society Series D-The Statistician	218	218	218
Journal Of Visual Languages And Computing		250	250
Land Use Policy			416
Landscape And Urban Planning			987
Lecture Notes In Computer Sciences	168	168	266
Papers In Regional Science			252
Pattern Recognition Letters			1604
Photogrammetric Engineering And Remote Sensing	1355	1355	1355
Professional Geographer		1291	1291
Progress In Human Geography	1572	1572	1572
Regional Science And Urban Economics			444
Regional Studies	1406	1406	1406
Remote Sensing Of Environment			1289
Risk Analysis			1171
Technometrics		476	476
Transactions Of The Institute Of British Geographers	667	667	667
Transportation Research Part C-Emerging Technologies			188
Urban Studies		1930	1930
Visual Computer		324	324

Appendix D. List of Journals and their contribution (in no. of papers) to 5-X co-authorship networks series

JOURNAL TITLE	5-4	5-3	5-2
Agriculture Ecosystems & Environment			1297
Annals Of The Association Of American Geographers		1183	1183
Applied Geography		463	463
Area			1094
Cartographic Journal		352	352
Cartography And Geographic Information Systems		56	56
Computer Journal			715
Computers & Geosciences	1223	1223	1223
Computers & Graphics			562
Computers Environment And Urban Systems	169	169	169
Earth Surface Processes And Landforms	914	914	914
Environment And Planning A	2180	2180	2180
Environment And Planning B-Planning & Design	1019	1019	1019
Environmental And Ecological Statistics			186
Environmental Management			1075
Fuzzy Sets And Systems		3393	3393
Geoderma			1137
Geographical Analysis			267
Geographical Journal			1198
Geography		1240	1240
Geoinformatica	118	118	118
Hydrological Processes			1435
International Journal Of Geographical Information Science	626	626	626
International Journal Of Remote Sensing	3015	3015	3015
Journal Of Geography In Higher Education			477
Journal Of Visual Languages And Computing		250	250
Lecture Notes In Computer Sciences	266	266	266
Photogrammetric Engineering And Remote Sensing	1355	1355	1355
Professional Geographer		1291	1291
Progress In Human Geography	1572	1572	1572
Transactions Of The Institute Of British Geographers		667	667
Transportation Research Part C-Emerging Technologies			188

Appendix E. Top IJGIS Authors and their non-included publication forums

Personal Web pages last visited on the 24th October 2006.

M. Goodchild	P.A. Burrough	M.J Egenhofer	R.T Aangeenbrug (1935-2003)	P. Fisher	D. Martin	N. Stuart	A.U. Frank	G.B.M. Heuvelink
http://www.geog.ucsb.edu/~good/#pubs	http://www.geo.uu.nl/phpscripts/staffpages/personal/personal.php?id=Burro101&menu=&infoid=1&groupid=0;	http://www.spatial.maine.edu/~max/pubs_RJ.html;		http://www.soi.city.ac.uk/~pff1/	http://www.geog.soton.ac.uk/users/martindj/davehome/publicn.htm	http://www.geos.ed.ac.uk/people/publications.html?indv=20	http://www.geoinfo.tuwien.ac.at/publications/index.php?by_author:Frank_Andrew_U.&#Journals	http://www.metis.wur.nl/result.cfm?r=result&i=publications&m=alphabetic&med=26157&year=&year2=
Cartographic Perspectives	Environmental and Ecological Studies	Informática Pública		IEEE Society and Technology	British Medical Journal	GeoInfoSystems	Journal of the Urban and Regional Information Systems Association	Transactions in GIS
Cartographica	Landscape Ecology	Spatial Cognition and Computation	v	Landscape Ecology	Government and Policy	Jounral of Biogeography	Spatial Cognition and Computation	v
GeoInfoSystems	Landschap	Transactions in GIS	x	x	Journal of Health Services Research and Policy	Journal of Applied Meteorology and Climatology	x	x
Geospatial Solutions	Transactions in GIS	x	x	x	Papers of the Regional Science Association	Journal of Belizean Affairs	x	x
International Journal of Applied Earth Observation and Geoinformation	x	x	x	x	Population Trends	Transactions in GIS	x	x
ISUMA: Canadian Journal of Policy Research	x	x	x	x	Transactions in GIS	x	x	x

Appendix F. MakingAuthrosCollaboration Combination Java Method

```
/**
*To get the addresses to each author. It creates a vector of
*strings. Each string
* is composed by all authors of the paper paired with thier *respectivly
address.
*/
 public void makeAuthrosCollaborationsCombinations()
 {
       //How many one authors and c1 addresses per paper?
       int howmanyEqual=0, inTotal =0; //equal addresses and authors
       int howmanyMoreAuthorsOnlyOne=0;
       int howmanyMoreAuthorsOthers=0;
       int  howmanyMoreCountriesMoreThanOne=0;//more  addressee  than  authors.
Difference Only One
       int  howmanyMoreCountriesOnlyOne=0;  //more  addresses  than  authors.
Difference One.

       //To make the combinations
       for (int i = 0; i < authorsPapers.size();i++)
       {
        //Get number of authors per paper
        IntnumAuthors=
 ((String)authorsPapers.elementAt(i)).split(";").length;
        //Get C1 addreesses
        String[] c1s=((String)c1Papers.elementAt(i)).split(";");
 //Get the countries (tokens) for all C1 addresses at the
 paper.
                                                         String[]c1CountryTk=
((String)c1CountriesPapers.elementAt(i)).split(";");
              int numCountries = c1CountryTk .length; //number of countries in
a paper C1 addresses
 //Get the paper's year
 Integer currentYear = (Integer)yearPapers.elementAt(i);
 //START DOING THE COMBINATIONS.
        if  (numAuthors==numCountries)  //1.  authors  number  ==  number  of  C1
addresses so 1-1 relation
        {
              //Do the pairing becasue is 1-1 relationship
              for (int j=0; j< numCountries-1; j++)
              {
               String country1 = c1CountryTk[j];
authorCountryTimesVec.add(country1+"\t"+currentYear.toString());   //add   one
country appearance
                           for (int k=j+1; k< numCountries;k++)
                           {
                                 String country2 = (c1CountryTk[k]);

       pairedAuthorsAddresses.add(country1+"\t"+country2+"\t"+currentYear);
                           }
        }
                     authorCountryTimesVec.add(c1CountryTk[numCountries-
1]+"\t"+currentYear.toString()); //add the last country appearance
                     howmanyEqual++;
        }
        else if (numCountries>numAuthors)//2. If there are more addresses than
authors and the diff is one.
        {
        //First For-Loop to solve the combination for the first position at
the String-Array

       authorCountryTimesVec.add(c1CountryTk[0]+"\t"+currentYear.toString());
//add the first country appearance
                           int difference = numCountries-numAuthors;
                           for (int k =difference+1; k< numCountries;k++)
```

```
                              {

pairedAuthorsAddresses.add(c1CountryTk[0]+"\t"+c1CountryTk[k]+"\t"+currentYea
r);
                              }
                              //To make the remaining combinations.
                              for (int l = difference+1; l< numCountries-1;l++)
                              {

       authorCountryTimesVec.add(c1CountryTk[l]+"\t"+currentYear.toString());
//add the group of countries after the difference appearance
                               for (int t= l+1; t< numCountries;t++)
                                     {

pairedAuthorsAddresses.add(c1CountryTk[l]+"\t"+c1CountryTk[t]+"\t"+currentYea
r);
                               }
                              }
                              if(difference==1)
                              {
                                     howmanyMoreCountriesOnlyOne++;
                              }
                        else
                        {
                                     howmanyMoreCountriesMoreThanOne++;
                        }
                              //add the last one
                              authorCountryTimesVec.add(c1CountryTk[numCountries-
1]+"\t"+currentYear.toString()); //add the first country appeareance
                }
                else //3. if there are more authors than addresses.
                {
                              String country1=null;
                              if  (numCountries==1)//3.1  If  there  one  addresses,
Assign it to all combinations.
                              {
                                     country1 = c1CountryTk[0];
                               for (int j=0; j< numAuthors-1; j++)
                                     {

authorCountryTimesVec.add(country1+"\t"+currentYear.toString());
                                      for (int k=j+1; k< numAuthors;k++)
                                            {

       pairedAuthorsAddresses.add(country1+"\t"+country1+"\t"+currentYear);
                                            }
                                     }
                               howmanyMoreAuthorsOnlyOne++;

authorCountryTimesVec.add(country1+"\t"+currentYear.toString());//last one
                              }
                              else//3.2   To   the   remaining   cases,   make   the
combinations among the existing addresses.
                              {
                                     for (int j=0; j< numCountries-1; j++)
                                     {

       authorCountryTimesVec.add(c1CountryTk[j]+"\t"+currentYear.toString());
                                            country1 = c1CountryTk[j];
                                            for (int k=j+1; k< numCountries;k++)
                                            {
                                             String country2 = (c1CountryTk[k]);

pairedAuthorsAddresses.add(country1+"\t"+country2+"\t"+currentYear);
                                            }
                                     }
```

```
      authorCountryTimesVec.add(c1CountryTk[numCountries-
1]+"\t"+currentYear.toString());
                                  howmanyMoreAuthorsOthers++;
                             }

                 }
 }//End-Most External For
                                   inTotal                          =
howmanyEqual+howmanyMoreCountriesOnlyOne+howmanyMoreCountriesMoreThanOne+howm
anyMoreAuthorsOnlyOne+howmanyMoreAuthorsOthers;
 System.out.println("Total = "+ inTotal );
 System.out.println("Countries Equal Authors = "+howmanyEqual);
                                            System.out.println("MoreCountries
"+(howmanyMoreCountriesMoreThanOne+howmanyMoreCountriesOnlyOne));
 System.out.println("\tDifference One= "+howmanyMoreCountriesOnlyOne);
 System.out.println("\tMore Than One= "+howmanyMoreCountriesMoreThanOne);
              System.out.println("Total            More             Authors
"+(howmanyMoreAuthorsOnlyOne+howmanyMoreAuthorsOthers));
 System.out.println("\tOnly One Address= "+howmanyMoreAuthorsOnlyOne);
 System.out.println("\tMore Than One= "+howmanyMoreAuthorsOthers);
 System.out.println("=========================================");
//sNA.printVector(authorCountryTimesVec,2,"Papers-Authors   Country");   //To
print country number of papers.
//printCombinations();
 }//End-Method
```

Appendix G. Complete list of author-collaborations per country, sorted by number of collaborations.

	Country	92	Country	92-96	Country	97-02	Country	92-02
1	United States	800	United States	2960	United States	5423	United States	8383
2	England	338	England	1417	England	2445	England	3862
3	France	160	India	948	Canada	969	Canada	1529
4	Canada	99	France	563	France	834	India	1504
5	Australia	97	Canada	560	Italy	775	France	1397
6	India	80	Italy	489	Germany	585	Italy	1264
7	Italy	73	Australia	332	India	556	Australia	859
8	Netherlands	67	Netherlands	249	Australia	527	Germany	812
9	Germany	47	Germany	227	Netherlands	494	Netherlands	743
10	Spain	36	Spain	227	China	492	Spain	629
11	Sweden	29	Scotland	174	Spain	402	China	556
12	Scotland	19	Greece	142	Japan	376	Japan	497
13	Japan	18	Japan	121	Scotland	293	Scotland	467
14	New Zealand	15	Wales	118	Belgium	262	Israel	361
15	Russia	15	Israel	114	Israel	247	Greece	321
16	Wales	13	Russia	105	Brazil	190	Belgium	317
17	China	11	New Zealand	95	New Zealand	180	Wales	292
18	Argentina	6	Sweden	89	Greece	179	New Zealand	275
19	Belgium	6	China	64	Wales	174	Brazil	251
20	Czech Republic	6	Brazil	61	Sweden	119	Sweden	208
21	Finland	6	Belgium	55	Switzerland	117	Russia	190
22	Saudi Arabia	6	Switzerland	53	Taiwan	109	Switzerland	170
23	Brazil	5	Turkey	50	Portugal	106	Taiwan	145
24	Israel	5	Norway	40	Mexico	100	Portugal	117
25	Denmark	4	Taiwan	36	Russia	85	Turkey	113
26	Greece	4	Hungary	33	Finland	77	Norway	110
27	Norway	4	Egypt	31	Norway	70	Mexico	108
28	Switzerland	4	Argentina	30	Austria	63	Finland	103
29	Austria	3	Austria	28	Turkey	63	Austria	91
30	Ireland	3	North Ireland	28	Denmark	59	Denmark	82
31	Thailand	3	Finland	26	South Korea	48	North Ireland	72
32	Egypt	2	Denmark	23	Botswana	45	Egypt	61
33	Hungary	2	Singapore	20	North Ireland	44	Singapore	58
34	Nigeria	2	South Africa	18	Ireland	39	Argentina	55
35	Taiwan	2	Saudi Arabia	16	Singapore	38	South Korea	55
36	Trinid & Tobago	2	Cameroon	12	South Africa	34	South Africa	52
37	Yugoslavia	2	Czech Republic	12	Egypt	30	Ireland	48
38	Costa Rica	1	Portugal	11	Argentina	25	Botswana	45
39	Morocco	1	Nigeria	10	Poland	24	Hungary	38
40	Albania	0	Ireland	9	Chile	23	Czech Republic	32
41	Algeria	0	Chile	8	Thailand	23	Chile	31
42	Bahrain	0	Mexico	8	Czech Republic	20	Poland	30
43	Bangladesh	0	Bulgaria	7	Estonia	20	Thailand	29
44	Belarus	0	Morocco	7	Kenya	20	Cameroon	25
45	Bermuda	0	South Korea	7	French Guiana	14	Estonia	25
46	Botswana	0	Philippines	6	Senegal	14	Kenya	24
47	Bulgaria	0	Poland	6	Algeria	13	Saudi Arabia	24
48	Cameroon	0	Romania	6	Cameroon	13	Malaysia	18
49	Chile	0	Thailand	6	Malaysia	13	Nigeria	18

50	Colombia	0	Ukraine	6	Slovakia	12	Senegal	16
51	Cote Ivoire	0	Albania	5	Costa Rica	11	Slovakia	15
52	Croatia	0	Estonia	5	Ethiopia	11	Algeria	14
53	Cuba	0	kuwait	5	French Polynesia	11	Costa Rica	14
54	Cyprus	0	Malaysia	5	Venezuela	11	French Guiana	14
55	Estonia	0	Kenya	4	Syria	10	Ukraine	14
56	Ethiopia	0	Trinid & Tobago	4	Iceland	9	French Polynesia	13
57	French Guiana	0	Yugoslavia	4	Tunisia	9	Venezuela	13
58	French Polynesia	0	Costa Rica	3	Nigeria	8	Albania	11
59	Georgia	0	Slovakia	3	Saudi Arabia	8	Ethiopia	11
60	Ghana	0	Croatia	2	Ukraine	8	Iceland	11
61	Guatemala	0	French Polynesia	2	Colombia	7	Syria	10
62	Honduras	0	Iceland	2	Albania	6	Morocco	9
63	Iceland	0	Jordan	2	Luxembourg	6	Philippines	9
64	Indonesia	0	Lithuania	2	Oman	6	Tunisia	9
65	Iran	0	Malawi	2	Bermuda	5	kuwait	8
66	Jordan	0	Nepal	2	Hungary	5	Romania	8
67	Kenya	0	Senegal	2	Cote Ivoire	4	Bulgaria	7
68	kuwait	0	Slovenia	2	Cyprus	4	Colombia	7
69	Laos	0	Venezuela	2	Honduras	4	Trinid & Tobago	7
70	Lebanon	0	Algeria	1	Indonesia	4	Luxembourg	6
71	Lithuania	0	Bahrain	1	Laos	4	Oman	6
72	Luxembourg	0	Bangladesh	1	Lebanon	4	Bermuda	5
73	Malawi	0	Indonesia	1	Namibia	4	Indonesia	5
74	Malaysia	0	Sudan	1	Guatemala	3	Nepal	5
75	Mexico	0	Belarus	0	Iran	3	Cote Ivoire	4
76	Namibia	0	Bermuda	0	kuwait	3	Cyprus	4
77	Nepal	0	Botswana	0	Nepal	3	Honduras	4
78	North Ireland	0	Colombia	0	Philippines	3	Laos	4
79	Oman	0	Cote Ivoire	0	Trinid & Tobago	3	Lebanon	4
80	Philippines	0	Cuba	0	U arab emirates	3	Namibia	4
81	Poland	0	Cyprus	0	Belarus	2	Slovenia	4
82	Portugal	0	Ethiopia	0	Georgia	2	Yugoslavia	4
83	Romania	0	French Guiana	0	Ghana	2	Guatemala	3
84	Senegal	0	Georgia	0	Morocco	2	Iran	3
85	Singapore	0	Ghana	0	Romania	2	U arab emirates	3
86	Slovakia	0	Guatemala	0	Slovenia	2	Bangladesh	2
87	Slovenia	0	Honduras	0	Zambia	2	Belarus	2
88	South Africa	0	Iran	0	Zimbabwe	2	Croatia	2
89	South Korea	0	Laos	0	Bangladesh	1	Georgia	2
90	Sudan	0	Lebanon	0	Cuba	1	Ghana	2
91	Syria	0	Luxembourg	0	Uganda	1	Jordan	2
92	Tunisia	0	Namibia	0	Bahrain	0	Lithuania	2
93	Turkey	0	Oman	0	Bulgaria	0	Malawi	2
94	U arab emirates	0	Syria	0	Croatia	0	Zambia	2
95	Uganda	0	Tunisia	0	Jordan	0	Zimbabwe	2
96	Ukraine	0	U arab emirates	0	Lithuania	0	Bahrain	1
97	Venezuela	0	Uganda	0	Malawi	0	Cuba	1
98	Zambia	0	Zambia	0	Sudan	0	Sudan	1
99	Zimbabwe	0	Zimbabwe	0	Yugoslavia	0	Uganda	1

Appendix H. Complete list of country-collaborations per country, sorted by number of papers.

	Country	92	Country	92-96	Country	97-02	Country	92-02
1	England	35	United States	205	United States	403	United States	608
2	United States	32	England	160	England	318	England	478
3	Canada	21	Canada	68	Canada	143	Canada	211
4	France	12	France	53	France	101	France	154
5	Germany	12	Scotland	48	Germany	97	Germany	140
6	Netherlands	12	Italy	45	Scotland	91	Scotland	139
7	Italy	11	Germany	43	Italy	88	Italy	133
8	Australia	8	Wales	38	Netherlands	87	Netherlands	123
9	Wales	6	Australia	37	Australia	67	Australia	104
10	India	3	Netherlands	36	China	66	Wales	102
11	Ireland	3	Greece	26	Wales	64	China	78
12	New Zealand	3	Spain	21	Brazil	51	Spain	68
13	Scotland	3	Israel	19	Belgium	50	Belgium	65
14	Spain	3	Russia	19	Japan	48	Brazil	62
15	Austria	2	India	16	Spain	47	Japan	58
16	Czech Republic	2	Belgium	15	New Zealand	36	Greece	56
17	Denmark	2	China	12	Israel	33	Israel	52
18	Finland	2	Brazil	11	Greece	30	New Zealand	45
19	Hungary	2	Japan	10	India	25	India	41
20	Sweden	2	Sweden	10	Sweden	25	Russia	39
21	Switzerland	2	Hungary	9	Switzerland	24	Sweden	35
22	Argentina	1	New Zealand	9	Denmark	21	Switzerland	30
23	Brazil	1	Argentina	6	Norway	21	Norway	27
24	China	1	Ireland	6	North Ireland	20	Denmark	26
25	Costa Rica	1	North Ireland	6	Russia	20	North Ireland	26
26	Greece	1	Norway	6	Finland	18	Taiwan	24
27	Israel	1	Switzerland	6	Taiwan	18	Finland	20
28	Morocco	1	Taiwan	6	Singapore	14	Austria	18
29	Russia	1	Austria	5	Austria	13	Singapore	18
30	Thailand	1	Czech Republic	5	South Korea	12	Ireland	15
31	Trinid & Tobago	1	Denmark	5	Kenya	11	Kenya	14
32	Albania	0	Albania	4	Portugal	10	Hungary	13
33	Algeria	0	Romania	4	Ireland	9	Portugal	13
34	Bahrain	0	Saudi Arabia	4	Mexico	9	South Korea	13
35	Bangladesh	0	Singapore	4	Estonia	8	Mexico	12
36	Belarus	0	Kenya	3	Poland	8	Argentina	10
37	Belgium	0	Mexico	3	South Africa	8	South Africa	10
38	Bermuda	0	Portugal	3	Turkey	8	Turkey	10
39	Botswana	0	Chile	2	Chile	6	Estonia	9
40	Bulgaria	0	Finland	2	Malaysia	6	Poland	9
41	Cameroon	0	kuwait	2	Senegal	5	Chile	8
42	Chile	0	Lithuania	2	Thailand	5	Czech Republic	8
43	Colombia	0	Morocco	2	Ukraine	5	Malaysia	7
44	Cote Ivoire	0	Philippines	2	Argentina	4	Thailand	7
45	Croatia	0	South Africa	2	Bermuda	4	Ukraine	7
46	Cuba	0	Thailand	2	Botswana	4	Albania	5
47	Cyprus	0	Turkey	2	Cameroon	4	kuwait	5
48	Egypt	0	Ukraine	2	Colombia	4	Philippines	5
49	Estonia	0	Algeria	1	Hungary	4	Romania	5
50	Ethiopia	0	Bahrain	1	Slovakia	4	Saudi Arabia	5

51	French Guiana	0	Bangladesh	1	Costa Rica	3	Senegal	5
52	French Polynesia	0	Bulgaria	1	Cyprus	3	Slovakia	5
53	Georgia	0	Costa Rica	1	Czech Republic	3	Bermuda	4
54	Ghana	0	Egypt	1	Egypt	3	Botswana	4
55	Guatemala	0	Estonia	1	Ethiopia	3	Cameroon	4
56	Honduras	0	Iceland	1	French Guiana	3	Colombia	4
57	Iceland	0	Indonesia	1	French Polynesia	3	Costa Rica	4
58	Indonesia	0	Malawi	1	Iceland	3	Egypt	4
59	Iran	0	Malaysia	1	kuwait	3	Iceland	4
60	Japan	0	Nepal	1	Namibia	3	Cyprus	3
61	Jordan	0	Poland	1	Oman	3	Ethiopia	3
62	Kenya	0	Slovakia	1	Philippines	3	French Guiana	3
63	kuwait	0	South Korea	1	Tunisia	3	French Polynesia	3
64	Laos	0	Sudan	1	Indonesia	2	Indonesia	3
65	Lebanon	0	Trinid & Tobago	1	Laos	2	Morocco	3
66	Lithuania	0	Venezuela	1	Nepal	2	Namibia	3
67	Luxembourg	0	Belarus	0	Syria	2	Nepal	3
68	Malawi	0	Bermuda	0	Venezuela	2	Oman	3
69	Malaysia	0	Botswana	0	Albania	1	Tunisia	3
70	Mexico	0	Cameroon	0	Algeria	1	Venezuela	3
71	Namibia	0	Colombia	0	Bangladesh	1	Algeria	2
72	Nepal	0	Cote Ivoire	0	Belarus	1	Bangladesh	2
73	Nigeria	0	Croatia	0	Cote Ivoire	1	Laos	2
74	North Ireland	0	Cuba	0	Cuba	1	Lithuania	2
75	Norway	0	Cyprus	0	Ghana	1	Syria	2
76	Oman	0	Ethiopia	0	Guatemala	1	Trinid & Tobago	2
77	Philippines	0	French Guiana	0	Honduras	1	Bahrain	1
78	Poland	0	French Polynesia	0	Iran	1	Belarus	1
79	Portugal	0	Georgia	0	Lebanon	1	Bulgaria	1
80	Romania	0	Ghana	0	Luxembourg	1	Cote Ivoire	1
81	Saudi Arabia	0	Guatemala	0	Morocco	1	Cuba	1
82	Senegal	0	Honduras	0	Nigeria	1	Ghana	1
83	Singapore	0	Iran	0	Romania	1	Guatemala	1
84	Slovakia	0	Jordan	0	Saudi Arabia	1	Honduras	1
85	Slovenia	0	Laos	0	Trinid & Tobago	1	Iran	1
86	South Africa	0	Lebanon	0	U arab emirates	1	Lebanon	1
87	South Korea	0	Luxembourg	0	Uganda	1	Luxembourg	1
88	Sudan	0	Namibia	0	Zambia	1	Malawi	1
89	Syria	0	Nigeria	0	Bahrain	0	Nigeria	1
90	Taiwan	0	Oman	0	Bulgaria	0	Sudan	1
91	Tunisia	0	Senegal	0	Croatia	0	U arab emirates	1
92	Turkey	0	Slovenia	0	Georgia	0	Uganda	1
93	U arab emirates	0	Syria	0	Jordan	0	Zambia	1
94	Uganda	0	Tunisia	0	Lithuania	0	Croatia	0
95	Ukraine	0	U arab emirates	0	Malawi	0	Georgia	0
96	Venezuela	0	Uganda	0	Slovenia	0	Jordan	0
97	Yugoslavia	0	Yugoslavia	0	Sudan	0	Slovenia	0
98	Zambia	0	Zambia	0	Yugoslavia	0	Yugoslavia	0
99	Zimbabwe	0	Zimbabwe	0	Zimbabwe	0	Zimbabwe	0

Appendix I. Complete list of country-collaboration Salton Measure (SM) and Collaboration Exclusivity Index (CEI) between 1992-1996

Country 1	Country 2	No country papers C1	No country papers C2	CEI	SM
Greece	Scotland	26	48	20%	42%
England	Wales	160	38	13%	32%
England	United States	160	205	10%	20%
Canada	United States	68	205	8%	19%
England	Scotland	160	48	8%	18%
Australia	England	37	160	7%	18%
Greece	Russia	26	19	9%	18%
Spain	Taiwan	21	6	7%	18%
Brazil	United States	11	205	4%	17%
Italy	Spain	45	21	8%	16%
Israel	United States	19	205	4%	14%
France	United States	53	205	6%	14%
Canada	Netherlands	68	36	7%	14%
England	Ireland	160	6	2%	13%
Japan	Argentina	10	6	6%	13%
Australia	United States	37	205	5%	13%
France	Germany	53	43	6%	13%
Italy	United States	45	205	5%	12%
Canada	England	68	160	6%	12%
Northern Ireland	Scotland	6	48	4%	12%
Taiwan	United States	6	205	2%	11%
France	Argentina	53	6	3%	11%
Belgium	France	15	53	4%	11%
England	Italy	160	45	4%	11%
Germany	Russia	43	19	5%	10%
Germany	Hungary	43	9	4%	10%
Switzerland	Canada	6	68	3%	10%
Australia	China	37	12	4%	9%
Netherlands	United States	36	205	3%	9%
France	New Zealand	53	9	3%	9%
Spain	United States	21	205	3%	9%
Japan	United States	10	205	2%	9%
Sweden	United States	10	205	2%	9%
China	Germany	12	43	4%	9%
Belgium	New Zealand	15	9	4%	9%
Hungary	India	9	16	4%	8%
France	Italy	53	45	4%	8%
Canada	Hungary	68	9	3%	8%
Russia	United States	19	205	2%	8%
England	India	160	16	2%	8%
England	New Zealand	160	9	2%	8%
Israel	Hungary	19	9	4%	8%
Russia	Hungary	19	9	4%	8%
Netherlands	Germany	36	43	4%	8%
Spain	Hungary	21	9	3%	7%
Russia	Japan	19	10	3%	7%
Belgium	United States	15	205	2%	7%
Canada	China	68	12	3%	7%

Germany	Israel	43	19	3%	7%
India	United States	16	205	2%	7%
Germany	Italy	43	45	3%	7%
Australia	North Ireland	37	6	2%	7%
North Ireland	Wales	6	38	2%	7%
Russia	Scotland	19	48	3%	7%
Wales	Norway	38	6	2%	7%
Italy	Scotland	45	48	3%	6%
Northern Ireland	England	6	160	1%	6%
Germany	United States	43	205	2%	6%
Scotland	Spain	48	21	3%	6%
Canada	Belgium	68	15	2%	6%
Belgium	England	15	160	2%	6%
Italy	Ireland	45	6	2%	6%
Switzerland	Italy	6	45	2%	6%
England	Germany	160	43	2%	6%
Spain	France	21	53	3%	6%
Russia	India	19	16	3%	6%
Norway	United States	6	205	1%	6%
Switzerland	United States	6	205	1%	6%
France	Norway	53	6	2%	6%
France	Switzerland	53	6	2%	6%
Australia	New Zealand	37	9	2%	5%
England	France	160	53	2%	5%
New Zealand	Wales	9	38	2%	5%
Israel	Russia	19	19	3%	5%
Belgium	Greece	15	26	2%	5%
England	Sweden	160	10	1%	5%
Italy	Netherlands	45	36	2%	5%
Norway	Canada	6	68	1%	5%
Germany	Japan	43	10	2%	5%
Sweden	Germany	10	43	2%	5%
Australia	Scotland	37	48	2%	5%
France	Netherlands	53	36	2%	5%
China	England	12	160	1%	5%
Brazil	Italy	11	45	2%	4%
Wales	France	38	53	2%	4%
India	Netherlands	16	36	2%	4%
Scotland	China	48	12	2%	4%
Australia	India	37	16	2%	4%
New Zealand	Canada	9	68	1%	4%
United States	China	205	12	1%	4%
France	China	53	12	2%	4%
England	Netherlands	160	36	2%	4%
Canada	Sweden	68	10	1%	4%
Japan	Canada	10	68	1%	4%
Israel	Netherlands	19	36	2%	4%
Israel	Australia	19	37	2%	4%
India	Italy	16	45	2%	4%
Wales	Russia	38	19	2%	4%
Canada	Germany	68	43	2%	4%
Brazil	Canada	11	68	1%	4%
Spain	Netherlands	21	36	2%	4%
Canada	Italy	68	45	2%	4%

Australia	Spain	37	21	2%	4%
Greece	Netherlands	26	36	2%	3%
Norway	England	6	160	1%	3%
Israel	France	19	53	1%	3%
England	Greece	160	26	1%	3%
India	Canada	16	68	1%	3%
Germany	Greece	43	26	1%	3%
Ireland	United States	6	205	0%	3%
Israel	Canada	19	68	1%	3%
Netherlands	Australia	36	37	1%	3%
Greece	United States	26	205	1%	3%
Wales	Australia	38	37	1%	3%
Scotland	Netherlands	48	36	1%	2%
England	Brazil	160	11	1%	2%
Wales	Scotland	38	48	1%	2%
United States	Hungary	205	9	0%	2%
United States	Wales	205	38	1%	2%
England	Israel	160	19	1%	2%
Canada	Scotland	68	48	1%	2%
Spain	England	21	160	1%	2%
Canada	France	68	53	1%	2%
Scotland	United States	48	205	0%	1%

Appendix J. Complete list of strengthened country-collaboration from 1992-1996 to 1997-2002.

country1	country2	1992-1996 SM	1997-2002 SM	Increment in SM of
United States	China	4%	17%	13.1%
Wales	Scotland	2%	9%	6.8%
England	Scotland	18%	24%	5.8%
Canada	France	2%	7%	5.8%
North Ireland	England	6%	11%	4.8%
Australia	New Zealand	5%	10%	4.7%
Japan	United States	9%	12%	3.4%
Netherlands	Germany	8%	11%	3.3%
Germany	United States	6%	10%	3.2%
India	United States	7%	10%	3.0%
Switzerland	Italy	6%	9%	2.6%
Israel	Russia	5%	8%	2.5%
Netherlands	Australia	3%	5%	2.5%
France	Italy	8%	11%	2.4%
Ireland	United States	3%	5%	2.1%
Scotland	United States	1%	3%	2.1%
Canada	United States	19%	21%	1.8%
Canada	Scotland	2%	4%	1.8%
Israel	Canada	3%	4%	1.6%
Spain	England	2%	3%	1.5%
England	Netherlands	4%	5%	1.5%
Belgium	United States	7%	8%	1.2%
England	Israel	2%	3%	1.1%
Brazil	Canada	4%	5%	1.0%
Scotland	Netherlands	2%	3%	1.0%
France	Norway	6%	7%	0.9%
Australia	India	4%	5%	0.8%
Germany	Italy	7%	8%	0.8%
Germany	Greece	3%	4%	0.7%
Australia	Scotland	5%	5%	0.4%
Israel	France	3%	3%	0.3%
Netherlands	United States	9%	10%	0.3%
New Zealand	Canada	4%	4%	0.1%
Belgium	Greece	5%	5%	0.1%
Germany	Israel	7%	7%	0.1%

Appendix K. Complete list of new country- collaboration partnerships during 1997-2002.

country1	country2	1997-2002 SM	country1	country2	1997-2002 SM
Malaysia	Singapore	22%	Sweden	Singapore	5%
Estonia	Finland	17%	Greece	Spain	5%
Singapore	India	16%	Chile	Wales	5%
Japan	Turkey	15%	Sweden	Wales	5%
Norway	Russia	15%	Mexico	United States	5%
Chile	South Africa	14%	China	Sweden	5%
Finland	Sweden	14%	England	South Korea	5%
Kenya	South Africa	11%	Japan	New Zealand	5%
Germany	Switzerland	10%	Italy	Denmark	5%
New Zealand	United States	9%	Spain	Portugal	5%
Belgium	Italy	9%	Malaysia	England	5%
Belgium	Portugal	9%	Switzerland	North Ireland	5%
Norway	Sweden	9%	Singapore	England	4%
Netherlands	China	8%	China	Greece	4%
Russia	Poland	8%	New Zealand	Singapore	4%
Estonia	Norway	8%	North Ireland	France	4%
Finland	Italy	8%	Portugal	Brazil	4%
North Ireland	Ireland	7%	Wales	South Africa	4%
Germany	Spain	7%	Sweden	Denmark	4%
Mexico	Denmark	7%	Denmark	United States	4%
Norway	Ireland	7%	South Korea	United States	4%
Wales	South Korea	7%	Belgium	Germany	4%
Estonia	Sweden	7%	Brazil	Germany	4%
Australia	South Korea	7%	Belgium	Kenya	4%
Poland	France	7%	Japan	Spain	4%
Brazil	France	7%	Scotland	India	4%
Israel	Sweden	7%	Chile	Germany	4%
Austria	Wales	7%	China	Ireland	4%
England	Switzerland	7%	Israel	Finland	4%
Germany	Denmark	7%	France	Chile	4%
Germany	Norway	7%	Brazil	Japan	4%
Netherlands	Kenya	6%	England	Poland	4%
Singapore	Taiwan	6%	Austria	Belgium	4%
Scotland	Sweden	6%	Italy	Greece	4%
Brazil	England	6%	Japan	Singapore	4%
Brazil	North Ireland	6%	Germany	Wales	4%
Belgium	Denmark	6%	Netherlands	Poland	4%
Poland	Israel	6%	Canada	Russia	4%
Belgium	Netherlands	6%	England	Mexico	4%
Kenya	United States	6%	Switzerland	Greece	4%
Turkey	England	6%	England	Denmark	4%
Netherlands	Austria	6%	Canada	Spain	4%

Canada	Taiwan	6%	China	South Korea	4%
Greece	Portugal	6%	Australia	Japan	4%
Brazil	Chile	6%	Estonia	France	4%
England	Japan	6%	South Africa	France	4%
Wales	Brazil	4%	Australia	Canada	2%
Japan	Finland	3%	Chile	United States	2%
Austria	Australia	3%	Malaysia	United States	2%
England	Kenya	3%	India	Germany	2%
Italy	Portugal	3%	England	South Africa	2%
Mexico	France	3%	Brazil	Belgium	2%
Scotland	Portugal	3%	Canada	Denmark	2%
Brazil	Finland	3%	Wales	Spain	2%
Australia	Singapore	3%	Greece	France	2%
North Ireland	Spain	3%	New Zealand	Netherlands	2%
Belgium	Russia	3%	Japan	China	2%
Austria	England	3%	England	Portugal	2%
Italy	Japan	3%	Wales	Belgium	2%
Brazil	Denmark	3%	Estonia	United States	2%
Canada	Estonia	3%	Poland	United States	2%
Spain	Sweden	3%	United States	South Africa	2%
Israel	New Zealand	3%	United States	Turkey	2%
India	Japan	3%	Belgium	China	2%
Germany	Austria	3%	Belgium	Australia	2%
Mexico	Canada	3%	Germany	New Zealand	2%
Austria	United States	3%	Sweden	Canada	2%
Austria	France	3%	United States	Portugal	2%
Russia	China	3%	Japan	Netherlands	2%
Australia	Norway	3%	Netherlands	Brazil	2%
Denmark	Australia	3%	Scotland	Brazil	1%
England	Finland	3%	Wales	Italy	1%
Greece	Brazil	3%	United States	Singapore	1%
Netherlands	Finland	3%	Taiwan	England	1%
Kenya	Canada	3%	China	Italy	1%
China	Switzerland	3%	Australia	Germany	1%
Russia	England	3%	Scotland	Germany	1%

Belgium	Israel	2%
Australia	Sweden	2%
South Korea	Canada	2%
Netherlands	North Ireland	2%
Finland	United States	2%
France	Finland	2%
Canada	Austria	2%
North Ireland	Germany	2%
North Ireland	United States	2%
France	Denmark	2%
Switzerland	Scotland	2%
Canada	Wales	2%
Scotland	France	2%
Spain	Belgium	2%
New Zealand	China	2%