A COMPLEXITY APPROACH TO EVALUATING NATIONAL SCIENTIFIC SYSTEMS THROUGH INTERNATIONAL SCIENTIFIC COLLABORATIONS

by

Ryan J. Zelnio

A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
The Requirements for the Degree
of
Doctor of Philosophy
Public Policy

Committee:

_________________________________ David Hart, Chair

_________________________________ Robert Axtell

_________________________________ Christopher Hill

_________________________________ Connie McNeely

Loet Leydesdorff, External Reader

_________________________________ James P. Pfiffner, Program Director

_________________________________ Edward Rhodes, Dean

Date: ___________________________ Spring Semester 2013
George Mason University
Fairfax, VA

A Complexity Approach to Evaluating National Scientific Systems through International Scientific Collaborations

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Mason University

by

Ryan J. Zelnio
Master of Science
University of North Dakota, 2004
Bachelor of Science
Marycrest International University, 1998

Director: David Hart, Professor
Department of Public Policy

Spring Semester 2013
George Mason University
Fairfax, VA

DEDICATION

This is dedicated to my loving wife Kelly, who put up with and supported me throughout this long ordeal and to my children Ariana and Madoc who are my inspiration for all things.

ACKNOWLEDGEMENTS

I would like to thank the many friends, relatives, and supporters who have made this happen. My loving wife, Kelly, my parents, Tom and Sharon Zelnio, and my in-laws Jack and Beverly Jones, as well as the rest of my family, provided me strong support and encouragement throughout my research. My fellow SPPers Dana Dolan, Yu Jin Jung, Ward Kay, Kadri Kallas and Fangmeng Tian for being sounding boards through my doctoral studies. A special thanks also to my Chair, Dr. David Hart for reading through a number of revisions and guiding me as I struggled through numerous issues issues and encouraging me to complete this project. Finally, thanks go out to members of the BAT development group at the Naval Surface Warfare Center – Dahlgren for supporting me while I finished my dissertation

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

ABSTRACT

A COMPLEXITY APPROACH TO EVALUATING NATIONAL SCIENTIFIC SYSTEMS THROUGH INTERNATIONAL SCIENTIFIC COLLABORATIONS

Ryan J. Zelnio, Ph.D.

George Mason University, 2013

Dissertation Director: Dr. David Hart

This dissertation seeks to contribute to a fuller understanding of how international scientific collaboration has affected national scientific systems. It does this by developing three methodological approaches grounded in social complexity theory and applying them to the evaluation of national scientific systems. The first methodology identifies the global core-periphery structure of science at the scientific disciplinary level. The second methodology creates a multitheoretic, multilevel model for studying the evolution of international collaborations. The third methodology develops a framework that relies on identifying scientific topics with a lack of international collaboration to identify areas in which a country has emerging capabilities that may give it an advantage in the global arena. Each of these methodologies are applied to a variety of case studies and can be used in conjunction to obtain a fuller understanding that could aid the governance and management of national and international investments in science.

CHAPTER ONE: INTRODUCTION

International scientific cooperation (ISC) grew exponentially over the past two decades (Glänzel 2001). This cooperation results in the transfer of knowledge between participants as well as the creation of new knowledge. The growth in international cooperation has had profound effects on the conduct of research both internal and external to national borders. This change in conduct in turn has implications for the effective functioning of national systems supporting science. Yet, the mechanisms and processes that underlie the growth in cooperation are poorly understood (Katz and Hicks 1997, Wagner-Döbler 2001, Wagner and Leydesdorff 2005). A fuller understanding of the factors contributing to this growth could aid the governance and management of national and international investments in science.

The factors affecting growth in ISC can be divided into two groups, those internal and those external to science (Wagner and Leydesdorff 2005, Drori et al 2003, Hwang 2008). Factors internal to science include both the conduct of science and the social structure of the scientific community. External factors are those coming from social and political environments in which scientists reside. Confounding our understanding of ISC, the two types of factors have an effect on each other. To unravel the relationship between these factors and the growth of ISC, this dissertation proposes utilizing the lens of social complexity theory. The central thesis of this social complexity theory was encapsulated in the title of Thomas Schelling's seminal work "Micromotives and Macrobehavior", which the decisions and actions of individuals have significant consequences for the group in which they reside. This dissertation applies this theory by examining the decision of scientists to collaborate with another scientist residing in another country having a dramatic impact on the structure of the nation scientific system of the countries in which they and their collaborator reside.

This introduction is split into three sections: 1) a brief literature review on international scientific collaboration; 2) methodological considerations; and 3) an overview of the dissertation

Literature Review of International Scientific Collaboration

The factors affecting growth in ISC can be divided into two explanatory themes, those internal and those external to science (Schott 1993, Wagner and Leydesdorff 2006, Dosi et al 2003, Hwang 2008). This literature review examines these two macro themes separately.

Factors internal to the conduct of science

The earlier of the two explanatory themes is that factors internal to the conduct of science are largely responsible for the increase in scientific cooperation (Merton 1968, 1973; Zuckerman 1989; Ben-David 1990; and others). This line of inquiry has been explored particularly in the "sociology of science" scholarship.

This explanatory theme captures the benefits associated with ISC that are internal to the conduct of science: access to more skills, knowledge, and techniques; transferring of knowledge; the desire of researchers to increase their visibility and recognition; intellectual companionship; and a clash of views that may lead to new insights (Katz and Martin, 1997). Thus, individuals within the scientific community are rewarded with greater status, more promotions, and increased funding should they conform to these mores of scientific culture and increase collaborative activities (Whitley 1984, Drori et al 2003).

Internal to science is a clear hierarchy among scientists, from Nobel Prize winners to graduate students and a defined path of how one is promoted in this hierarchy (Merton 1968). Scientists at the top of this hierarchy are accorded the greatest status and have higher visibility. Senior scientists are sought after as collaborators by junior scientists as a way to increase their own visibility and add additional credibility to the junior scientist's own research (Merton 1968; Price and Gursey 1976). This effect is commonly referred to as the "Matthew effect," after the biblical passage in the Gospel of Matthew describing the effect that "the rich get richer and the poor get poorer." This hierarchy also fits within the concept of "preferential attachment," the process in which the attractiveness of a scientist for collaboration is directly proportional to how well known the scientist is. Price and Gursey (1976) found that this effect leads to senior scientists gaining even greater prominence as they increase their reputation through an increasing number of collaborations while junior scientists' reputation grows at a much smaller rate as they are partially eclipsed by the more prominent collaborator.

This hierarchy also extends to the international level. There is a core-periphery structure associated with the uneven distribution of science globally (Schott 1993, Kim 2006, Hwang 2008). Within this core-periphery structure, scientists from core countries are given much greater status and scientists in the periphery countries are encouraged to collaborate with scientists located within the core to have access to the latest equipment and theories.

The changing conduct of science is causing scientists to be more dependent on collaboration within their own scientific networks (Schott 1992, Glänzel 2001, Wagner 2008, Hwang 2008). This change is driven by the increased specialization within the sciences, the growing interdisciplinary nature of science, and the increased mobility of scientists (Schott 1992, Hwang 2008).

The various internal factors are institutionalized in part through the formation of professional societies (Drori et al. 2003). These societies are self-organizing with membership and advancement reflective of the hierarchy observed by Merton (Beaver and Rosen 1978). The rapid growth of professional societies in core countries (Drori et al. 2003) is evidence of the fragmentation of science resulting from increases in specialization and interdisciplinary science.

There is a strong correlation between the expansion of professional science within core nations and the increase of professional science in international non-governmental organizations (INGOs) (Drori et al. 2003). The formation rate of INGOs has increased exponentially, practically doubling every 10 years since the 1950s (Drori et al. 2003). These INGOs reflect a culture that encourages and promotes scientific cooperation regardless of political borders. These INGOs have expanded into the role of providing assistance to developing countries on matters relating to scientific public goods by promoting cooperation among members with scientists in the developing world.

Another explanatory theme internal to science emphasizes the nature or culture of exploration among scientists that is self-reinforcing (Drori, et al. 2003). It takes the view that science is a public good to be shared globally among universities and laboratories in order to accelerate the rate of discovery. The success of science endogenously supplies its research agenda and creates incentives for autonomous expansion. Where resources are scarce, the scientific community seeks to internally prioritize where to allocate these resources. Examples of this would be the decadal surveys performed by the National Academies and the peer-reviewed grant process used by the National Science Foundation. The decadal surveys are conducted among scientists to find out what the major scientific questions are that need to be attacked within a specific discipline. One intention of the authors of these surveys is that they should be used by national policy-makers to decide where to allocate resources for science.

"Big science", or scientific endeavors that requires large-scale projects for progress, was initially viewed as a major driver in the growth of ISC (see Gallison and Heavly 1992). The rising cost of instrumentation for certain fields such as space science and particle physics necessitated that countries share the cost of pursuing such endeavors. An empirical study by Wagner (2008) comparing the growth in ISC between a "big science", astrophysics, and "small sciences" such as soil science and mathematics logic found that ISC, as measured by coauthored publications, grew at a rate of 138% in astrophysics between 1990 and 2000, while the small sciences of mathematical logic and soil science grew at the rates of 333% and 323% respectively during the same period. While the growth rate of these small sciences is particularly impressive, one's conclusions drawn from these observations about growth rates should be tempered with the fact that astrophysics had already established ISC as a norm, while the others were more recently internationalized. In 2000, 47% of all papers in astrophysics were internationally coauthored, compared with 38% in mathematical logic and 33% in soil sciences.

The analysis by Wagner highlights that increases in ISC differ among fields of science. While "big science" has a pressing need for ISC due to the costs involved, other fields such as the soil sciences cited above require access to certain regions or virology which may require access to specific populations. Cooperation becomes necessary to access these local resources. Still others, such as theoretical physics, do not require any sharing of physical resources and collaboration is mainly driven by a desire of scientists to "consult with others". Through studying collaborative patterns among differing branches of science, one may differentiate the effects of factors that are internal to the conduct of science on the degree of ISC.

Factors external to the conduct of science

External to the conduct of science is a social context that provides motivations that are external to the scientific context and include: the rising cost of doing research; the declining cost of travel and communication leading to a growing availability and easy access to researchers; and political factors encouraging greater collaboration across borders (Katz and Martin 1997).

One may hypothesize that growth in information and communication technologies (ICT) alone helps explain the sudden increase in ISC seen over the past two decades. Wagner (2006) argues that scientists enabled through advances in ICT have empowered this factor recently by lowering barriers associated with pursuing international scientific collaboration (ISCs) in "small" science. While growth in ICT is definitely an enabling factor, work by Laudel (2001) has shown that the majority of ISC begin face-to-face. Thus, ICT can be seen as a necessary but not sufficient part of the growth of collaborations.

A competing explanatory theme encompassing factors external to science is that the conduct of science is a product of the society in which it is embedded (for examples see Callon et. al. 1986; Latour and Woolgar 1986). This theme suggests that social factors external to science, such as national and institutional politics, heavily control the conduct of scientists. Within this framework, scientists are seen as entrepreneurs seeking to exploit opportunities within the social context external to science to promote their own agendas. If national policies state that funds will be provided for research involving international cooperation, scientists will thus include international teams with little thought given to the scientific merit of adding such members. Indeed, diplomacy has been found to be a significant driver in the increase of ISC (Oldham 2005, Skolnikoff 2001).

The condition of a country's scientific infrastructure plays a role in its propensity to cooperate. Luukenoon, Persson, and Siverston (1992) found that "[t]he less developed the scientific infrastructure of a given country, the higher the tendency for international coauthorship collaboration." The increased specialization of science has also had an effect on scientists from countries with small scientific output who are forced to look outside their borders for collaborative partners. Thus ISC may also be viewed to be a good way of "enhancing domestic scientific capabilities" (Oldham 2005). This effect has been tempered by the increased transaction costs of such collaborations.

The transaction costs associated with ISC occur at both the institutional and national level. At the institutional level these costs have additional financial burden, they take more time to accomplish, require increased administration, and have significant problems in reconciling institutional systems such as financial, management, and property rights (Katz and Martin 1997). At the national level, there are also political impediments in advanced countries for gathering support for ISC. These include the belief that other countries are free-riding, that ISC will be giving away know-how to other countries, and that science will be subordinated to strategic and political ends (Wagner 1997). Thus countries may erect barriers to collaboration which include export controls, classification, and visa restrictions.

There are additional costs imposed in the short-term to developing countries that accompany an increase in ISC (Schofer, et. al. 2000, Drori et. al. 2003). These costs arise from the fact that increases in the scientific workforce and scientific research activities are heavily correlated with women's rights, human rights, consumer's rights, gay and lesbian rights, and environmental rights. While the increase in all these rights is a social benefit to a developing country, they create a corresponding economic cost that is burdensome for developing countries to bear.

The transaction costs associated with collaboration may be mitigated by foreign policies to address either social factors (to address creation of knowledge that would benefit global public goods, such as health, agriculture, or environmental issues) or for political factors (such as foreign aid programs or to promote better relations between countries) (Oldham 2005). Additionally, shared geography, culture, and languages may also play a role in lowering transaction costs (Oldham 2005, Ponds et al 2006).

Methodological Considerations

The bulk of the empirical section of this dissertation proposal is a detailed coauthorship analysis of articles cataloged in Thomson Reuter's ISI Web of Knowledge (WoK). WoK is one of the most comprehensive databases of peer-reviewed journals covering over 256 subjects dating back 100 years. Coauthorship analysis (CA) is a widely used bibliometrics approach to understanding ISC. A paper is coauthored if it has more than one author, and is internationally coauthored if at least two authors list addresses from different countries. Smith (1958) was one of the first advocated the use coauthored papers to measure the possible increase in scientific collaboration and de Solla Price (1963) was one of the first to use direct bibliometrics measurement to support Smith's hypothesis. Since then a consensus has emerged that coauthorship is synonymous with collaboration with the one caveat being that there are cases of scientific fraud in which "honorary authorships" are given. (Katz and Martin 1997)

Issues in bibliometric data collection

CA is performed by looking at the author and address fields of papers. The address field describes both the physical location of researchers as well as their institutional affiliation. The basic units of analysis for collaborative studies using CA are author, institution, and country (Gauffriau et al. 2007). Some researchers (for example see Glänzel and Schubert 2005) also use the number of citations to provide some kind of weight to the importance of the collaboration.

Katz and Martin (1997) outlined several advantages for the use of bibliometrics approaches for CA. First, it is invariant and verifiable. Second, it is inexpensive and a practical means for quantifying collaborations. Third, the sample size is typically large and thus statistically significant. Lastly, it is an unobtrusive and non-reactive measurement.

There are some problems with the use address fields. Wagner and Leydesdorff (2005) found that ~5% of papers had more addresses than authors. This occurs because some authors claim to be from multiple institutes. Thus the number of authors gives a lower limit to the number of institutions that are represented through the collaboration.

Gauffiau et al. (2007) notes several issues with doing CA. The first is that within Thompson Scientific's large database of journal article, which is comprised of SCI (Science Citation Index), SSCI (Social Science Citation Index) and AHCI (Arts & Humanities Citation Index), about 20 million addresses are listed. This is in contrast to the estimated 50,000 different institutes that conduct research world-wide. One reason for this is that the coding for institutes may or may not contain the name of the institute itself, but may be just a department, university, or any combination of these three factors. This provides a significant hurdle in automating the coding of institutional affiliation of authors and requires intensive hand coding.

This problem is lessened somewhat for doing country coding of articles. While generally, countries can be deciphered unambiguously, they can appear under different names and are sometimes omitted and can only be inferred from their institutional affiliation or city.

There are three separate counting methods for determining the number of authors, institutions, and/or countries represented in any given publication (Gauffiau et al 2007). The first is whole counting which counts all unique authors, institutions, or countries contributing to a publication. The second is fractional counting which normalizes counting by assigning 1/n credit to each author, institute, or country where n is equal to the total number of participants. The third is first author or first address counting which provides a rank dependent accounting in which only the first listed address given is tabulated. This research is concerned primarily with whether a country participated in collaborative research, and not the degree of participation. Therefore whole counting will be used.

There is also an issue with which scientific disciplines can be studied using bibliometrics. At issue is whether or not the scientific discipline main discourse is held in peer reviewed publications that are indexed by one of the two main databases, ISI's Web of Knowledge or Elsevier's Scopus. These two databases are the most complete collections of peer-reviewed journals across many disciplines, though there are others that specialize in gathering so called "gray literature" such as conference proceedings like IEEE and Google scholar.

Stefaniak (2001) examined its possibilities of CA for social science and found three major flaws in its use. The first is that social science has a tendency to publish in more types of literature, such as books and non-peer-reviewed journals, than other sciences. Secondly, as stated above, most citation databases do not include books. Lastly, unlike "hard" sciences, English is not the universal language and social scientists tend to publish in their own language. There is a heavy bias in most databases towards English-based journals and their coverage of other languages tends to be lacking. For this reason, my research will not examine social science disciplines.

Moed (2005) has done considerable work cataloguing ISI's coverage by discipline. The following table shows the results of his work. His work on coverage is based on how many of the works cited in papers are located within ISI's database. In general, coverage of theoretical and health-related sciences are excellent, applied research has good to moderate coverage and social science and humanities have poor coverage.

Table 1 ISI coverage of disciplines (Moed 2005)

Excellent (80%+)	Good (60%+)	Moderate(40%+)	Poor (<40%)
Molecular biology and biochemistry	Applied physics and chemistry	Engineering	Other social sciences

Biological sciences primarily relating to humans	Biological sciences primarily related to animals and plants	Mathematics	Humanities and arts
Clinical Medicine	Psychology & psychiatry	Economics	
Physics & astronomy	Other social sciences primarily related to medicine and health		
Chemistry	Geosciences		

Creating and analyzing the empirical ISC networks

Gathering the empirical data is perhaps one of the most challenging aspects of this research proposal. The following sections outline the general strategy for accomplishing this task.

Collecting the data

The first decision to be made in the data collection is determining which data to collect. It has been known for quite some time that the level of collaboration varies by discipline (for example see Luukonen et. al. 1992). Therefore, any analysis of variation needs to sample a sufficiently broad swath of enough disciplines over a sufficiently long enough time period to determine if there are trends in international collaboration that are internal to science.

Disciplines shall be selected based on WoK discipline classification (a full listing of disciplines categorized by ISI is in the appendix). WoK uses Bradford’s Second Law loosely to determine which journals compromise a discipline. This law simply states that a relatively small number of core journals publish the bulk of significant scientific research. There are exponential diminishing returns in extending search for relevant articles outside of the core journal set as the journals tend to be more strongly correlated with other disciplines. WoK uses a citation analysis to determine the core set of journals for any discipline and supplements this quantitative analysis with a qualitative assessment that includes whether the journal adheres to basic journal standards, editorial content, and international diversity. (Thomson)

The next decision is how many years of data to collect. Prior research done by Wagner and Leydesdorf (2005) compared ISC network structures from 1990 and 2000 in six different disciplines and later updated with 2005 data (Leydesdorf and Wagner, 2008). While doing snapshots at these time periods is useful in demonstrating that the pattern of collaboration has changed, it does not provide enough detail to shed light into how the patterns of collaboration have evolved. For this, more fine-grained annual analysis may be required. Thus for the chapter 2 and 3, data will be collected for all years between 1983 and 2008. This data range has the advantage of allowing a swath of data spanning a quarter of a century and covering several significant historical changes in detail including the collapse of the Soviet Union and the internet boom.

For research into emerging trends, a much shorter time period is required. Previous studies of emergence have looked at clustering time slices of 3 years (Glänzel and Thijs 2012) and 6 years (Upham and Small 2010) without much discussion as to what is the ideal time span. For better statistical analysis, the QEHS described in chapter 4 has erred on the upper part of the range and chosen 6 years as the time span.

Aggregating the data

There are multiple ways in which one may aggregate the data, each of which can provide valuable insights. The data may be aggregated either at the institutional level or at the national level. The institutional level is useful in that it provides insights into the types of institutions that participate within a country’s border. It can be used to illuminate the differences in collaboration strategies for universities, research institutes and private companies for various nations. The national view though provides insight into the ramification of national priorities. One of the central question in this research is how is international collaboration affecting the conduct of science. For a large part, science is funded at the national level. Thus, this research will be looking at countries as the primary unit of analysis.

Next, as discussed in the previous step, I'll be collecting data over a continuous 25-year time period for chapters 2 and 3 (in chapter 4, the data is not aggregated beyond the single year). Whether or not this is aggregated can provide different insights. My preliminary work, based on an examination of nanoscience and nanotechnology, suggested the aggregation of results into 3 time periods: Cold War (1983-1992), Internet Boom (1993-2000), and modern era (2001-2008). This had the advantage of creating a large enough sample size to have significant results for certain analysis such as using power law analysis to test for small world properties, which requires a minimum of two orders of magnitude to be significant. However, this size of aggregation may have hidden underlying patterns which may shed insight into how nations change position in the core-periphery structure. For instance, South Korea had no collaboration and only 0.9% share of publication in the Cold War Era, which then jumped to 67 joint papers with 5.1% of share in the internet boom and 326 joint papers and 8.6% share in the modern era. The movement was dramatic but the fidelity of the data was not sufficient for gaining understanding of when Korea began to substantially increase its international presence. This proves problematic in the second part of the analysis when I concentrate on examining country performance.

In addition to aggregating this data, certain statistics need to be gathered during this phase for later analysis. These statistics include number of papers authored, number of papers authored with international collaborators, and the total and average citations for papers that were authored both with and without collaboration.

Overview of the Dissertation

This dissertation is organized as three empirical chapters that utilize different methodologies from social complexity theory to peel back and reveal different insights into evaluating and understanding of a country's national scientific system. The organization begins by looking at the global level, then narrows down to country-to-country analysis and then narrowing down to only considering the country in isolation. Each chapter is written as a publishable article intended for submission in a peer-reviewed journal.

Chapter 2 utilizes examines the properties of the power law structure of article outputs and degree centrality distribution of countries in five different fields (astronomy & astrophysics, energy & fuels, nanotechnology & nanosciences, nutrition, and oceanography). This applies the methodology detailed in "Power Laws and Non-Equilibrium Distributions of Complexity in the Social Sciences" (Cioffi-Revilla, forthcoming) to scientometrics to uncover a two-tiered power law structure that is used to determine a country's position in the core-periphery structure in the global scientific enterprise.

Chapter 3 applies the concept of multitheoretical, multilevel network analysis (Monge & Contractor, 2003) to understanding the evolution of inter-relationship between countries. Using the same five fields as chapter 2, it examines five countries' (Chile, China, France, South Korea and the United States) and the African continent's changing pattern of international scientific collaboration to identify how their scientific systems have evolved over time.

Chapter 4 attempts to tackle one of the core elements of social complexity theory, emergence. Whereas the previous two chapters uses measurements of international collaboration to ranking relationships between countries, this chapter exploits the lack of international collaboration and data mining to identify topics in which a country could develop new emerging scientific and technological capabilities that could give it an edge over other countries. This research develops a new methodology, called Quantitative Exploratory Horizon Scanning based on this principle and then applies it the United States as a case study.

Finally, Chapter 5 draws conclusions of how this research can be used by policy makers and outlines further research. It also discusses how agent-based modeling, another key methodology from social complexity theory, can be used for future research.

CHAPTER TWO: IDENTIFYING THE GLOBAL CORE-PERIPHERY STRUCTURE OF SCIENCE[1]

Ryan Zelnio[1]

[1] *ryjaz@yahoo.com*
George Mason University, Fairfax, VA, USA

Abstract
While there is a consensus that there is a core-periphery structure in the global scientific enterprise, there have not been many methodologies developed for identifying this structure. This paper develops a methodology by looking at the differences in the power-law structure of article outputs and degree centrality distributions of countries. This methodology is applied to five different scientific fields: astronomy & astrophysics, energy & fuels, nanotechnology & nanosciences, nutrition, and oceanography. This methodology uncovers a two-tiered power law structure that exists in all examined fields. The core-periphery structure that is unique to each field is characterized by the core's size, minimum degree and exponent of its power law distribution. Stark differences are identified between technology and non-technology intensive scientific fields.

Keywords
Core-periphery structure; power law analysis; network centrality; global science

Introduction
There is a hierarchical, core-periphery structure in science that has dominated the international relations for scientific cooperation throughout the 20th century (Ben-David 1971, Traweek 1988, Schott 1998, Hwang 2008). Historically, the core structure was composed of the United States, Japan and the Europe Union (Ohmae 1985, Glänzel et al 2008). Socio-cultural elements such as nationality, colonial past, scientific heritage, and infrastructure reinforced the core-periphery structure (Oldham 2005, Hwang 2008). This structure predetermined the status of scientists and institutions creating a disadvantage for countries in the periphery to invest in the human capital necessary to advance their own institutions to be equal to those in the core (Schott 1998, Hwang 2008).

[1] This chapter has been published as: Zelnio, R., "Identifying the global core-periphery structure of science." Scientometrics, Vol.91, No. 2 (2012) 601-615

However, this triad and the corresponding core-periphery structure began to break down in the 21st century. This break down is due in part to the exponential growth over the past two decades in international scientific cooperation (ISC) (Glänzel 2001). This cooperation results in the rapid creation of new knowledge and transfer of knowledge across borders. This growth in international cooperation has a profound effect on the conduct of research both internal and external to national borders. These changes in conduct in turn have implications for the effective functioning of national systems supporting science (Hill 2007).

These changes began during the turn of the century and continue to this day. Even as these changes have taken place, there continues to be, as King (2004) put it, “a stark disparity between the first and second division in the scientific impact of nations”. This second division, the developing world, is increasingly becoming marginalized and exploited by those countries in the core (Lall 2001, Oldham 2005).

Yet, the mechanisms and processes that underlie the growth in cooperation at the national level is poorly understood (Katz and Hicks 1997, Wagner-Döbler 2001, Wagner and Leydesdorff 2005). While certain countries, mainly from Asia, have invested heavily in the sciences (Wagner & Leydesdorff 2005, Glänzel et al 2008, Hwang 2008, Leydesdorff & Wagner 2008) it is not well understood which of these investments have been successful in bridging the divide between the core and periphery. This research seeks to contribute to our understanding of the underlying structure of the core and periphery through quantitative analysis and understand how it evolves over time, with some countries entering and others exiting the core.

To gain new insights, this research will follow recent trends to examine scientific collaborative networks as collaborative networks are emergent, complex adaptive systems (Newman 2001, 2004, Barabasi et al 2002, Wagner and Leydesdorff 2005 among others). A characteristic common among emergent systems is that order appears to arise spontaneously from the local interaction of actors who are not necessarily aware of how their actions contribute to the larger order (Holland 1998). Building on this theoretical foundation, this research will extend the work of previous scholars by aggregating scientific collaborative networks to the nation-state level to perform a macro study on the evolution of countries to the core and periphery.

Data Collection and Methodology

Data was drawn from Thomson Reuter's ISI online bibliometrics database, Web of Knowledge (WoK) and the country names were standardized.[2] WoK is one of the most comprehensive databases of peer-reviewed journals covering over 256 subject areas dating back 100 years. Each subject area has a group of corresponding group of journals that provide a broad coverage on related topics. To provide a wide breadth for analysis, five subject areas were chosen: Astronomy and Astrophysics, Oceanography, Energy and Fuels, Nanosciences and Nanotechnology, and Nutrition. The first two subjects are chosen to be representative of the basic sciences; the next two for applied sciences and the last draws from medicine. Statistics for these fields are shown in Table 2.

Table 2 Subject Area Statistics[3]

Field	Coverage	Journals	Articles	Avg. Citation

[2]England, Wales, and Scotland are treated as separate entities in WoS and were combined into the United Kingdom. Additionally, East and West Germany were combined into a single entity prior to 1989.
[3] Nanotechnology & Nanoscience's coverage is shorter than others as the first journal classified in this subject area started at 1981. Additionally, 2008 data was not available at the time data collection occurred.

Astronomy	1979-2008	54	228796	24.3
Energy	1978-2008	74	88886	7.7
Nanotech	1981-2007	47	74755	7.5
Nutrition	1979-2008	69	95350	16.2
Oceanography	1979-2008	57	64764	18.8

To conduct this analysis, abstracts from all journals associated with the selected fields were downloaded and parsed into a database for analysis using customized software. Only research articles were analyzed, conference proceedings, reviews, editorials and letters were ignored. The number of research articles reported in Table 2 includes only articles with author address information included. With the exception of Energy and Fuels, less than 2.5% of articles did not have address information. For energy and fuels, slightly less than 12% of articles did not contain author information.[4]

For longitudinal panel analysis, publications were split into 3 major time periods: 1978-1992, 1993-2000, and 2001-2008. The first time period represents the global polarity in the world between the USSR and USA. The second time period was chosen as it is widely recognized that an explosive growth in international scientific publications followed the post-Soviet time period (Georghiou, 1998; Glänzel, 2001), and the last time period is the current state of field.

Bilateral ties are counted for any instance in which authors from two countries were in the same publication. For articles with authors from more than two countries, each country was counted as country-pairs. For example, for an article with authors from three countries, the article was counted three times, once as a collaboration between country A and country B, then between country B and country C, and finally between

[4] The cause for this discrepancy was not investigated. However, the author hypothesizes that part of this discrepancy may come from the fact that WoS includes a large number of professional journals.

country A and country C. This results in countries having more bilateral ties than the total number of articles. To compensate for this, the percentages of articles that are part of an international scientific collaboration (ISC) are calculated as 1 minus the ratio of number of national papers (with no international coauthors) and the total count of articles containing a least one author from the country. Network centrality measurements were calculated using UCINET with all edges symmetrized and weighted by the total count of articles.

Power Law Analysis Methodology

A persistent pattern associated with complex system is power laws. In bibliometrics, power laws are commonly seen in publication frequencies (Lotka 1924), citation frequencies (de Solla 1965), the degree of authors in coauthorship networks (Newman 2001, Jeong et al. 2001, Barabassi et al. 2002) and the degree of international coauthorship (Wagner and Leydesdorff 2005).

The majority of analysis used for this paper is based on log-CCDF (Complementary Cumulative Distribution Function) power law which takes the form (Cioffi-Revilla, forthcoming):

Equation 1Log-CCDF Power Law Function

$$\log[1-\Phi(x)] = a' - (b-1)\log x$$

which yields a C.D.F.

Equation 2Log-CCDF Power Law C.D.F.

$$\Phi(x) = 1 - ax^{1-b}$$

and a corresponding probability density function (PDF)

Equation 3Log-CCDF Power Law P.D.F.

$$p(x) = \frac{a(b-1)}{x^b}$$

This type of a power law is the same type seen in coauthorship networks and citation networks. In order to determine the power law coefficients, data will be fit into a log-log format seen in equation 1 and then OLS regression will be used to compute the coefficients. A combination of the t-statistic and standard error associated with the power law exponent as well as the R^2 will be used to determine the goodness-of-fit to the power law (Cioffi-Revilla, forthcoming).

Core-Periphery Analysis Methodology

Core-periphery structures are often discussed but lack a formal definition (Borgatti & Everret 1999). When discussing the world scientific system, the core is the area which produces the majority of new science and has a high inwardness, while the periphery consumes the knowledge (Schott 1993, 2001, Hwang 2008). Using this definition or ones similar to this, scientometrics has been used to identify the core through the study of citations, publications and coauthorship (Glänzel 2001, Glänzel et al 2008, Wagner & Leydesdorff 2005, Leydesdorff & Wagner 2008, Hwang 2005, 2008).

Barabasi et al. (2002) found that in some coauthorship networks, a two-tier structure in the degree distribution appears with a cross-over point that varies by discipline. Each of the tiers has a different power-law coefficient. Wagner & Leydesdorff (2005) found similar two-tier architecture in studying international coauthorship networks. This is two tier architecture was coined by Zhou and Mondragón's (2004) study on internet topology as the "rich-club phenomenon." Serrano (2008) extended this to weighted networks and found rich-clubs in air transportation, world trade and coauthorship networks.

This research uses the advances in detecting rich-club phenomenon detailed in Serrano (2008) to determine the core countries in the various fields under study and how they have changed over time. The methodology has been adapted into the following algorithm for computing the different power law coefficients and core-membership:

1. Compute the weighted degree of all nodes
2. Compute the log-CCDF as outlined in the power methodology section
3. Use piece-wise linear regression of log-degree and log-CCDF using a moving threshold to get a best fit for power law coefficients

In some cases piecewise linear regression found more than one best fit, in such cases the value which has the great discontinuity with the prior degree centrality was chosen. For example, in oceanography from 2001-2008, the fit for the core beginning at 425 degree and 652 where very close, however the prior for degree value in was 295 and 546 respectively. Thus the discontinuity is greater at 425, thus the four countries with a degree centrality between 425 and 652 were considered part of the core.

Analysis and Results

To grasp the complex underlying patterns associated with international scientific cooperation (ISC) at the macro level, the analysis is broken down into two components:

1. Field Analysis
2. Core-Periphery Analysis

Each analysis is done over three distinct time periods over five separate disciplines to insight into the evolving nature of ISC. The results of each of these analyses are interpreted independently within this section.

Field Analysis

Statistical Properties

An analysis of collaboration at the disciplinary level shows that collaboration has grown significantly for the selected five case studies over the past thirty years. The results of this analysis are shown in tables 3 and 4.

Table 3 Field Statistics

Field	Time Period	Total Records	Average Citation	Average Authors	Average Countries	Percent ISC paper
Astronomy & Astrophysics	pre-1992	68076	31.85	1.42	1.15	11.5%
	1993-2000	68026	28.66	2.09	1.42	28.5%

	2001-2008	92694	15.62	3.27	1.8	45.3%
Energy & Fuels	pre-1992	29670	7.98	1.08	1.02	1.4%
	1993-2000	15301	10.63	1.33	1.09	8.3%
	2001-2008	43915	6.52	1.64	1.16	14.1%
Nanotech & Nanosciences	pre-1992	6701	15.91	1.13	1.02	2.5%
	1993-2000	19216	10.23	1.48	1.12	11.5%
	2001-2007	48838	5.28	2.01	1.23	20.0%
Nutrition	pre-1992	31175	20.87	1.38	1.04	3.3%
	1993-2000	24758	20.09	1.83	1.12	9.9%
	2001-2008	39417	10.15	2.53	1.24	18.2%
Oceanography	pre-1992	17672	30.09	1.17	1.04	3.6%
	1993-2000	18733	23.9	1.53	1.16	13.2%
	2001-2008	28359	8.48	2.09	1.33	25.4%

Table 3 shows that there are stark differences in the composition of each field. Astronomy and Astrophysics is the most collaborative of the fields surveyed, while Energy & Fuels is the least collaborative. Nutrition is the second most collaborative field surveyed in terms of co-authors yet is fourth in terms of percentage of paper that are collaborated abroad and nearly tied with Nanotech & Nanosciences for average countries. This difference in growth of average authors and countries seen in nutrition shows evidence that increases in coauthorship does not lead directly to similar increases in international collaboration, though the two remain highly correlated. The extent of how collaboration has evolved in these fields can be seen in table 4.

Table 4 Field Growth Rates

Field	***Avg. Authors***		***Avg. Countries***		***Percent ISC***	
	1993-2000	**2001-2008**	**1993-2000**	**2001-2008**	**1993-2000**	**2001-2008**
Astronomy & Astrophysics	47.2%	56.5%	23.5%	26.8%	147.8%	58.9%
Energy & Fuels	23.1%	23.3%	6.9%	6.4%	492.9%	69.9%
Nanosciences & Nanotech	31.0%	35.8%	9.8%	9.8%	362.5%	74.4%
Nutrition	32.6%	38.3%	7.7%	10.7%	200.0%	83.8%
Oceanography	30.8%	36.6%	11.5%	14.7%	266.7%	92.4%

The growth in average authors is highly correlated with the pre-existing amount co-authorship. That is, the higher the average of number of authors, the higher the growth rate. This indicates there may exist some type of snowballing phenomena which reflects the changing structure of a field. Thus the more collaborative a field is in general, the more authors choose in the future to coauthor in greater team sizes.

The same pattern does not hold for the average countries collaborating on a paper. Astronomy & astrophysics continues to have the strongest growth rate, making it increasingly more global. Of interest though is the difference in growth between nutrition and oceanography. In the initial time period, pre-1992, these two fields had the same average countries. However, their growth rates diverged with oceanography becoming globalized at a rate much faster than nutrition.

The growth in the percentage of articles in a field that are internationally scientific collaborations (articles that had a minimum of at least 2 authors from different countries) grew rapidly in the 1990s. However, the rate of growth was not sustainable and dropped considerably in the most recent time period. The greatest growth was seen in the fields which had the lowest rate of collaboration to start with. The smallest growth came in astronomy & astrophysics, which was the field that already had the highest level of collaboration. The difference of growth rates between nutrition and oceanography followed a pattern similar to that seen the average countries.

Distribution Analysis

In addition to the mean averages and growth, one can look at the distribution of country collaborations. As previously noted, power laws have a long and varied association with bibliographic data. Due to this association, the distribution of how many countries collaborate on a given publication was tested to see if it fit to a power law.

Figure 1 shows the log-log plots of CCDF verse country representation in a given publication for each of the five fields studied with the goodness of fit to a power law in table 5.

Table 5 Power Law Goodness of Fit for Country Collaboration Distribution

Field	Time Period	b + 1	log a	R2	std err	t-stat
Astronomy & Astrophysics	pre-1992	3.728	-0.579	0.967	0.231	-16.14
	1993-2000	3.317	-0.097	0.951	0.195	-17.01
	2001-2008	3.241	0.414	0.865	0.286	-11.34
Energy & Fuels	pre-1992	2.955	-1.987	0.986	0.160	-18.45
	1993-2000	3.344	-1.149	0.986	0.182	-18.4
	2001-2008	3.489	-0.949	0.986	0.138	-25.26
Nanotech & Nanosciences	pre-1992	3.780	-1.631	0.995	0.196	-19.28
	1993-2000	4.503	-0.776	0.982	0.309	-14.56
	2001-2007	4.382	-0.452	0.987	0.188	-23.29
Nutrition	pre-1992	2.777	-1.734	0.957	0.241	-11.52
	1993-2000	2.648	-1.077	0.984	0.102	-25.99
	2001-2008	3.034	-0.460	0.939	0.214	-14.19
Oceanography	pre-1992	3.923	-1.409	0.992	0.199	-19.69
	1993-2000	3.452	-0.853	0.975	0.228	-15.14
	2001-2008	3.525	-0.342	0.990	0.111	-31.83

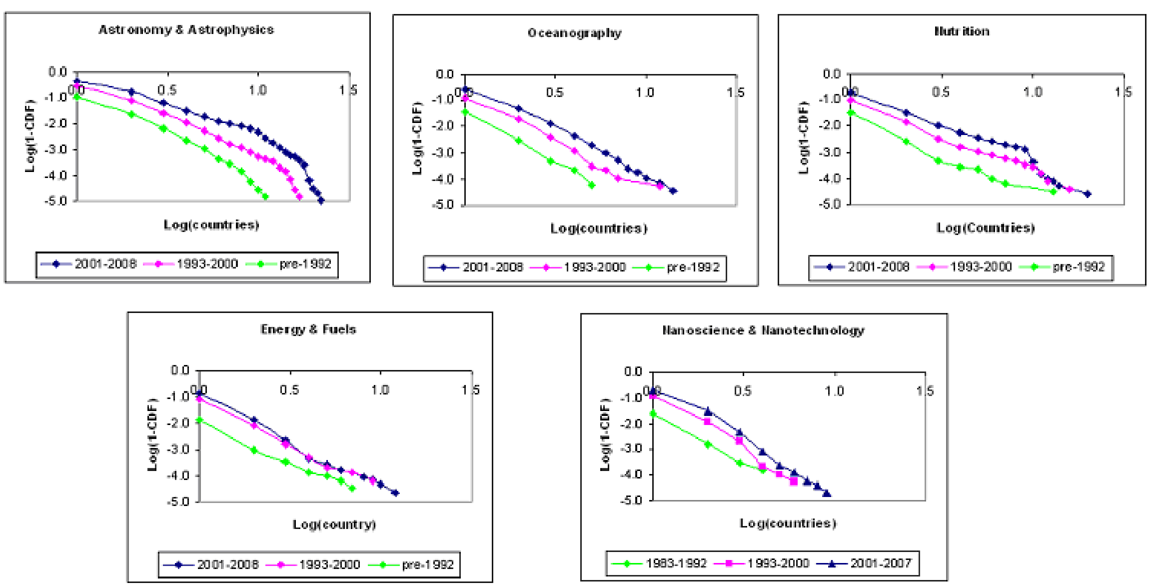


Figure 1 Distribution of Country Collaborations

The goodness-of-fit is strong for strong for all except the most collaborative of the fields, Astronomy and Astrophysics. A&A has a very strong hook at the end of the 2001-2008 data which starts at 19 countries collaborating. When this hook is removed, OLS regression shows a slope (b+1) of -2.61 with an intercept (log a) of 0.023 and the goodness-of-fit goes up considerably with a R2 of 0.954, standard deviation of 0.143 and t-statistic of -18.21. The hook appears in pre-1992 data starting at when 10 or more countries collaborate and in the 1993-2000 data when 15 more countries collaborate. A similar hook, though not as pronounced, shows in the 2001-2008 graph for nutrition, starting after log 1, or when more than 10 countries collaborate together. A likely explanation of this hook comes from the literature surrounding power-laws seen in social networks. It has been noted by Amaral et al. (2000) and Wagner and Leydesdorff (2005) that the cost of adding additional vertices to a network is a limiting factor, especially in collaborative networks. However, over time the hook begins at a later point suggests that these costs have been decreasing over time at a different rate for each field.

This analysis shows that the scale-free collaboration found by Newman (2001), Jeong et al.(2001), and Barabassi at al.(2002) at the individual scientist level, and scale-free collaboration of scientists collaborating with international scientists found by Wagner and Leydesdorff (2005) also applies at the macro level of countries collaborating together to publish articles. Just like the other studies, the power-law exponent varies by disciplines but falls in a similar range between 2.6 and 3.6, with the exception of nanosciences & nanotechnology (N&N). One of the reasons why N&N's slope is considerably steeper may lay in the fact the N&N is a much "younger" discipline with a smaller select number countries investing heavily in the discipline. This hypothesis will be explored later during the country-level analysis of this chapter.

Core-Periphery Analysis

Degree Distribution

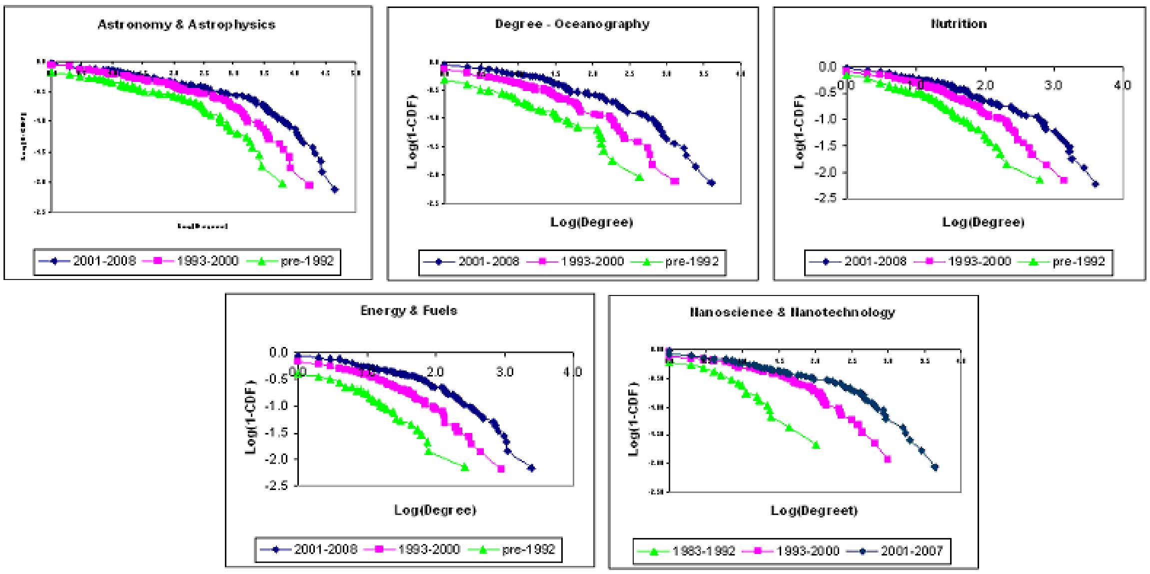


Figure 2 Distribution of Country's Degree Centrality

Figure 2 shows the initial graph of the distribution of degree centrality in log-log space. The significant shift to the right over time show that the density of cooperation among countries is growing across all disciplines. The majority of graphs also show a marked discontinuity in which the distribution bends towards a different slope. This bend is characteristic of the core-periphery structure previously discussed and is illustrated in figure 3.

Piece-wise linear regression shows two distinct structures lie within this distribution (see figure 3). An analysis shown in table 7 shows this structure permeating across all disciplines studied.

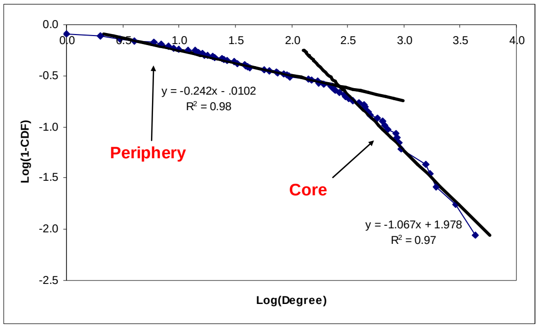


Figure 3 Nanotechnology and Nanosciences 2001-2007 Degree Centrality Distribution

The cohesiveness of the core varies by field and is measured through a combination of the core (power-law) coefficient, the minimum degree level of core countries, and the core size. The core coefficient is similar to a Gini coefficient seen in economic studies: the closer it comes to 0, the greater the disparity among members. Therefore, the greater the absolute value of the coefficient, the more cohesive the core. The inverse is also true: the greater the minimum degree of the core, the greater the density of cooperation within the core. The analysis in table 6 shows three distinct core-periphery patterns among the fields.

The first pattern is a highly dense and interconnected core and a largely disperse from the periphery as seen in Astronomy & Astrophysics. Over time, as the size of the core has grown so too has its density as reflected by the high minimum degree level seen in each time period. Additionally, the disparity between the core and periphery coefficients is suggestive that it is increasingly difficult for countries to travel from the periphery to the core in this discipline in the future.

The second pattern features a large but highly disconnected core membership. This is seen in the disciplinary fields of energy and nano, which are mostly loosely related to technological fields. This pattern suggests that countries are working largely independently of each other, more so in energy and fuel research than in nanotechnology.

The third pattern falls in between the first two and features a much smaller core of countries than the other disciplines but one that is highly dense and is reflected in oceanography and nutrition. This pattern has high core coefficients and a minimum degree ranging twice that seen in the second patter that suggests a high level of interconnectedness and more equity in the distribution of cooperation amongst core members.

Table 6 Degree Core Coefficients

	Astronomy			Nanotechnology			Energy			Nutrition			Oceanography		
	-'92	93'-00	01-'08	-'92	93'-00	01-'08	-'92	93'-00	01-'08	-'92	93'-00	01-'08	-'92	93'-00	01-'08
Core Coef.	-0.87	-1.01	-1.03	-0.94	-1.11	-1.07	-1.02	-1.15	-1.16	-1.10	-1.30	-1.42	-1.38	-1.23	-1.29
R-Sq	0.97	0.97	0.96	0.94	0.98	0.97	0.95	0.98	0.97	0.94	0.99	0.97	0.85	0.97	0.98
std error	0.03	0.04	0.04	0.13	0.04	0.03	0.08	0.04	0.05	0.11	0.04	0.06	0.26	0.06	0.04
t-stat	-25.68	-22.79	-24.53	-7.11	-31.05	-31.12	-12.96	-29.12	-24.72	-9.81	-33.92	-21.89	-5.42	-20.38	-30.46
Per. Coef.	-0.21	-0.23	-0.19	-0.57	-0.29	-0.24	-0.55	-0.44	-0.33	-0.56	-0.45	-0.39	-0.48	-0.42	-0.36
R-Sq	0.99	0.97	0.99	0.91	0.93	0.98	0.93	0.95	0.94	0.92	0.93	0.96	0.97	0.94	0.95
std error	0.00	0.01	0.00	0.06	0.02	0.01	0.04	0.02	0.01	0.03	0.02	0.01	0.02	0.02	0.01
t-stat	-48.78	-41.10	-71.52	-9.64	-18.01	-42.43	-14.25	-26.56	-26.84	-19.24	-24.56	-37.74	-23.82	-23.80	-32.41
Min. Degree	595	1147	2937	16	83	225	24	68	185	79	153	558	117	170	425
Core Size	13	20	26	7	19	29	10	19	23	9	15	18	7	14	17
Core Share	84.3%	89.2%	91.7%	82.1%	87.8%	94.7%	76.1%	78.7%	82.0%	75.0%	79.9%	76.1%	72.2%	80.2%	81.4%

Table 7 Article Core Coefficients

	Astronomy			Nanotechnology			Energy			Nutrition			Oceanography		
	-'92	93'-00	01-'08	-'92[5]	93'-00	01-'08	-'92	93'-00	01-'08	-'92	93'-00	01-'08	-'92	93'-00	01-'08
Core Coef.	-0.86	-0.88	-0.87	n/a	-0.86	-0.95	-0.81	-1.05	-1.20	-0.75	-0.99	-1.09	-0.76	-0.98	-1.15
R-Sq	0.95	0.97	0.97	n/a	0.99	0.98	0.94	0.94	0.95	0.95	0.95	0.95	0.95	0.96	0.95
std error	0.06	0.04	0.03	n/a	0.02	0.03	0.07	0.08	0.07	0.06	0.07	0.05	0.04	0.05	0.07
t-stat	-13.31	-22.88	-29.91	n/a	-44.06	-33.15	-11.37	-12.80	-17.67	-13.59	-14.50	-20.24	-16.92	-19.06	-16.58
Per. Coef.	-0.27	-0.24	-0.21	n/a	-0.24	-0.25	-0.41	-0.39	-0.32	-0.38	-0.39	-0.32	-0.37	-0.35	-0.32
R-Sq	0.95	0.97	0.99	n/a	0.94	0.97	0.89	0.94	0.94	0.95	0.95	0.96	0.97	0.98	0.98
std error	0.01	0.01	0.00	n/a	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
t-stat	-29.98	-41.69	-68.93	n/a	-22.95	-36.60	-21.57	-30.66	-32.92	-32.09	-34.72	-40.49	-32.47	-41.65	-50.87
Min. Article	505	859	1040	n/a	173	601	579	550	700	584	613	511	169	274	607
Core Size	12	20	29	n/a	22	21	10	13	19	11	12	23	16	17	18
Core Share	85.4%	89.3%	94.1%	n/a	91.7%	89.8%	83.0%	73.8%	80.6%	83.5%	77.3%	85.5%	92.40%	87.60%	84%

[5] There was no significant core-periphery structure during this time period.

Article Distribution

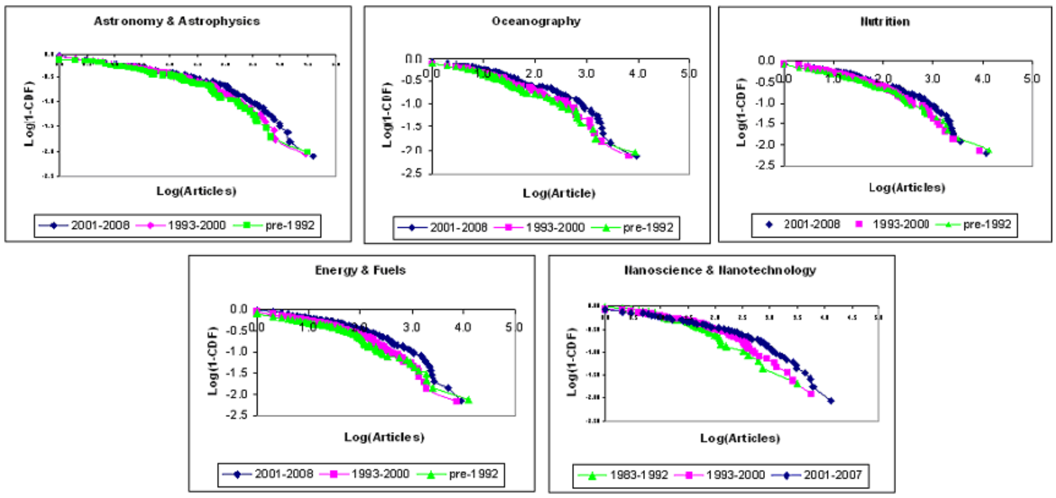


Figure 4 Distribution of Country's Article Output

Figure 4 shows the initial graph of the distribution of a country's article output in log-log space.The article output is the number of articles published in which at least one author was from a given country. Just as in the degree distribution, there is a strong 2-tier structure evident in the graph. However, in contrast to degree distribution seen in figure 2, there is not as pronounced growth (symbolized by a right shift in the graph) in overall article output. The analysis of the two-tiered structure is shown in table 7.

The power law exponent is significantly lower in both the article count's core and periphery structures than in the degree distribution. Thus, while overall output continues to grow and the core gets wider, the inequality of output between countries has decreased at a rate much slower than inequality of cooperation signifying that countries are depending more on cooperation for growth rather just outputting new articles.

In astronomy, while the core has grown over time, there has been little change in the core coefficient. A possible interpretation of this would be that the structure has reached some type of equilibrium. A possible explanation is that only astronomy has reached this equilibrium state would be that it has long had the highest amount of international cooperation and has reached some type of plateau.

In contrast, nanotechnology, the youngest of all the disciplines studied, went from having no discernable core-periphery structure in the first time period (the distribution fit closer to a single power law structure rather than a two-tier structure) to a large core in the second time period that then stabilized in the third time period with output become more equal in the core.

Energy and nutrition saw little growth in the minimum amount of articles published to enter the core while having a large increase in the core coefficient. A quick glance of the graph in figure 4 for these disciplines show a distinct break in the distribution illustrates this behaviour. One possible explanation could be that investment in core countries for these disciplines has stabilized.

Oceanography shows a similar break in the figure 4 but in this case there is significant increase in the minimum articles but little growth in the size of the core. This may be the case that countries interested in oceanography have increased their investments whereas other countries choose not to pursue investments, quite possibly due to the geographic nature of this discipline. Indeed, landlocked countries in Europe, notably Austria and Switzerland, which are part of the core in may other disciplines are not part of the core of oceanography.

Core-Periphery Membership

The core-periphery structure seen in the degree and article distributions has significant membership overlap though the rank order varies considerably. The majority of the differences of country membership occur with countries near the edge of the core-periphery structure. A good example of this can be seen in nutrition. In this field, India is publishing enough articles to be in the core of the article distribution but has very low degree centrality due to the fact that it has little international coauthorship (<7% prior to 2000 and 16% from 2001-2008). Switzerland on the other hand has the opposite pattern; it has a very low number of articles and doesn't enter the core in article counts prior to 2001 but has the highest level of internationally coauthored publications of any European country in all three time periods thus putting it squarely in the core in the degree centrality rankings. In India's case, you have a country working heavily in the area of nutrition science without accessing the global talent pool thus keeping them in the periphery. Whereas in Switzerland you have a country that is dependent on working with the core countries while not having enough of its own production to reach the critical threshold until the 2001-2008 time period. Thus it is for this reason that for a country to considered a core country within a discipline it must exist in both the degree and article cores. The result of this analysis is shown in table 8.

Table 8 Number of disciplines in which the country has core membership

Country	*pre-'92*	*93'-00*	*01-'08*
Canada	5	5	5
France	5	5	5
Germany	5	5	5
U.K.	5	5	5
USA	5	5	5
Japan	5	5	5
Italy	4	5	5
Spain	1	5	5

Country	*pre-'92*	*93'-00*	*01-'08*
Australia	3	4	5
Netherlands	2	4	5
Sweden	1	4	5
China	0	3	5
Russia	1	4	4
India	1	2	3
Switzerland	0	2	3
Denmark	1	1	3

Country	pre-'92	93'-00	01-'08
Belgium	0	1	3
Brazil	0	1	3
South Korea	0	1	3
Poland	0	2	2
Norway	1	1	2
Greece	0	0	2
Finland	0	0	2
Taiwan	0	0	2
Austria	0	2	1

Country	pre-'92	93'-00	01-'08
Mexico	0	1	1
Chile	0	1	1
Singapore	0	0	1
New Zealand	0	0	1
Turkey	0	0	1
Czech Republic	0	1	0
South Africa	1	0	0

The membership in the cores is similar to pattern observed by Glänzel (2001, 2008). The EU-US-Japan triad has dominated the core historically, with some exceptions. After the fall the USSR, Russia joined the core of the scientific community in all disciplines except nutrition. With the exception of New Zealand and Chile, most the other countries that have limited membership to the cores enter through investments in energy & fuels and nanotechnology, while those with membership in 3 cores also include astronomy and astrophysics. New Zealand's and Chile's core membership can be attributed primarily to geography as New Zealand is an island nation and Chile has many countries putting their southern observatories in the arid Atacama Desert. This research also confirms observation by Glänzel et al (2008) that China is integrating in the core of all sciences, though its membership in the non-technology focus disciplines of nutrition and oceanography is at the edge of the core.

Conclusion

In summary, even though there is a distinct and quantifiable core-periphery structure evident that varies across fields of science. The countries that tend to compose the majority of the core are similar with variations around the edges of the core. Originally comprised of a handful of countries in the core of any field of science, the core has grown to include 29 countries that maintain membership in at least one core. There are several implications to this phenomenon. For non-technology intensive fields, membership in the core increases the attractiveness of a country's scientific base. However, those countries left in the periphery are becoming increasingly isolated from the scientific elite. This isolation is growing as the gap widens in their relative decline in article output, degree centrality and citation impact. To understand how to overcome this isolation, a more detailed analysis of the collaborative behaviour at the bilateral level needs to be conducted.

Acknowledgements

An extended version of a paper presented at the 13th International Conference on Scientometrics and Informetrics, Durban (South Africa), 4–7 July 2011 (Zelnio, 2011).

The author would like to thank the conference attendees and the anonymous conference reviewers for their comments, many of which were incorporated.

References

Amaral, L.A.N. & J.M. Ottino (2004). Complex networks: Augmenting the framework for the study of complex systems. *Eur. Phys. J. B* 38, 147–162

Barabasi, A.L., Jeong, H.,N´eda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica A* 311, 590–614.

Ben-David J. (1971). *The Scientist's Role in Society*. Englewood Cliffs, NJ: Prentice-Hall.

Borgatti, S.P., and Everett, M.G. (1999). Models of Core/Periphery Structures. *Social Networks* 21, 375–395

Cioffi-Revilla, C. (eds.) (forthcoming). *Power Laws in the Social Sciences: Discovering Complexity and Non-Equilibrium Dynamics in the Social Universe*. Publication forthcoming.

De Solla, D.J. (1965) Networks of Scientific Papers. *Science* 149 (3683), 510-515

Gauffiau, M., P.O. Larsen, I. Maye, A. Roulin-Perriard, & M. Von Ins (2007). Publication, cooperation and productivity measures in scientific research. *Scientometrics*, Vol. 73(2), 175–214

Georghiou, L. (1998). Global cooperation in research. *Research Policy* 27, 611–626.

Glänzel, W. (2001). National characteristics in international scientific co-authorship relations. *Scientometrics*, Vol. 51(1), 69–115

Glänzel, W., and Schubert A. (2004). Analyzing scientific networks through co-authorship. In: Moed, H.F., Glänzel, W., and Schmoch, U. (2004) (eds.) *Handbook of quantitative science and technology research. The use of publication and patent statistis in studies of S&T systems* (pp. 257-276). Dordrecht (the Netherlands): Kluwer Academic Press

Glänzel, W., Debackere, K., and Mayer, M. (2008). 'Triad' or 'tetrad'? On Global changes in a dynamic world. *Scientometrics,* Vol. 74(1), 71-88

Hill, C.T. (2007). The Post-Scientific Society. *Issues in Science & Technology,* 24(1)

Holland, J. (1998). *Emergence: From Chaos To Order.* New York: Helix Books

Jeong, H., Neda,A., & Barabasi, A.L. (2001). Measuring preferential attachment for evolving networks. arXiv:cond-mat/0104131 v1, 7 April 2001.

Katz, J.S., & Hicks, D. (1997). How much is a collaboration worth? A calibrated bibliometric model. *Scientometrics* 40 (3), 541–554.

King, DA. (2004). The Scientific Impact of Nations. *Nature* V432

Lall, S. (2001). Competitiveness indices and developing countries: an economic evaluation of the Global Competitiveness Report, *World Development*, 29(9), 1501-1525.

Leydesdorff, L., & C.S. Wagner (2008). International Collaboration in Science and the Formation of a Core Group**.** *Journal of Informetrics* 2(4), 317-325

Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences* **16** (12), 317–324.

Merton, R.K. (1968). The Matthew Effect in Science The reward and communication systems of science are considered**,** Science 1968

Merton, R.K., (1973). *The Sociology of Science: Theoretical and Empirical Investigations.* Chicago: University of Chicago Press

Newman, M.E.J. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences* 98, 404–409.

Newman, M.E.J., (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences* 101 (Suppl. 1), 5200–5205.

Ohmae, K. (1985). *Triad Power: The Coming Shape of Global Competition*. New York: Free Press

Oldham, G. (2005). Policy Brief: International scientific collaboration: a quick guide, Science and Development Network, April 2005

Schott, T. (1993). World science: globalization of institutions and participation. *Science, Technology and Human Values*, 18, 196–208

Schott, T. (1998). Ties between center and periphery in the scientific world system. *Journal of World-Systems Research* 4, 112-144

Serrano, M.A. (2008). Rich-club vs rich-multipolarization phenomena in weighted networks. Physical Review E 78, 026101

Traweek, S. (1988). *Beamtimes and lifetimes: The world of high energy physicists*. Cambridge, MA:Harvard University Press.

Wagner C.S. (2005). Six case studies of international collaboration in science. *Scientometrics* 62 (1), 3-26

Wagner, CS (2008). *The New Invisible College: Science for Development*. Washington DC: Brookings Institution Press

Wagner C.S., Leydesdorff L, (2005). Network structure, self-organization, and the growth of international collaboration in science. *Research Policy* 34 (10), p.1608-1618

Wagner-Dőbler, R., (2001). Continuity and discontinuity of collaboration behaviour since 1800 – from a bibliometric point of view. *Scientometrics* 52, 503–517.

Zelnio, R. (2011), Structure of International Scientific Collaboration. In: E. Noyons, P. Ngulube, J. Leta (Eds), Proceedings of ISSI 2011 - The 13th International Conference on Scientometrics and Informetrics, Durban, South Africa, 4-7 July 2011, 898-913.

Zhou, S. and R. J. Mondragón (2004). The Rich-Club Phenomenon in the Internet Topology. *IEEE Communication Letters* 8(3), 180-1

CHAPTER THREE: METRICS FOR THE EVALUATION OF INTERNATIONAL SCIENTIFIC COLLABORATION IMPACT ON NATIONAL SCIENTIFIC SYSTEM

Abstract

Where science is happening has dramatically changed and no longer only concentrated in a few core countries. The impact of this shift is reflected in the rapid growth in international collaboration in the past 2 decades. This presents new challenges to policy makers to understand how to assess a country's scientific portfolio and position it within a global context. The policy goals vary dramatically. Developed countries may seek to maintain dominance or form strategic partnership, developing countries may be seeking pathways to strengthen their systems, and other organizations may be seeking to fulfill humanitarian causes of transferring knowledge from developed countries to developing countries. Built into this challenge is the need for metrics to evaluate progress towards meeting these goals. This paper lays out scientometric methodologies that may be used for evaluation. It then uses five case studies in which these methods are used in conjunction for evaluation. The first two case studies are an evaluation of the United States and France, both highly developed countries. The second two case studies are of Chile's and South Korea's rise from the periphery to the core in astronomy and nanotechnology respectively. The last case study is the interaction between the developed world and Africa in the field of Nutrition.

Introduction

Where science is happening has dramatically changed and no longer only concentrated in a few core countries (Glänzel et all 2008, Leydesdorff and Wagner 2009, Zelnio 2012). The impact of this shift is reflected in the rapid growth in international collaboration in the past 2 decades (Glänzel 2001, Wagner and Leydesdorff 2005). No longer is the leading edge of science being done in a handful of developed countries (United States, Western Europe and Japan); it is now spread globally (Glänzel et all 2008, Leydesdorff and Wagner 2009, Hill 2007, Zelnio 2012).

This presents new challenges to policy makers to understand how to position their country's scientific portfolios. The old paradigm of viewing national scientific systems in isolation has been tossed aside and the new paradigm in which policy makers must learn how best to position a country's system in the global web of science (Wagner 2008). Approaches for addressing this challenge vary depending on the current state of a nation's scientific infrastructure and the involvement of international NGOs such as the UNESCO, OECD, and the World Bank (Schott 2001, Wagner 2008). Built into this challenge is the need for metrics to evaluate their progress (Wagner 2008).

The objective of this paper is two-fold. First, it develops scientometric methodologies that may be used by policy makers to evaluate a country's international collaboration in terms of positioning and performance over time. Second, it demonstrates how these methodologies can be used in conjunction to evaluate the research portfolios of developing and developed countries. This is accomplished through five case studies. The first two case studies are an evaluation of the United States and France, both highly developed countries. The second two case studies are of Chile's and South Korea's rise from the periphery to the core in astronomy and nanotechnology respectively. The last case study is the interaction between the developed world and Africa in the field of Nutrition.

Data Collection

Data used is an expansion on the dataset used in Zelnio (2102) that includes all articles published between 1978 and 2008 in five subject areas: Astronomy and Astrophysics, Oceanography, Energy and Fuels, Nanosciences and Nanotechnology, and Nutrition. These subjects were chosen to provide broad coverage. The first two subjects are representative of the basic sciences; the next two are applied sciences and the last draws from medicine. Statistics for these fields are shown in Table 9.

Table 9 Subject Area Statistics[6]

Field	Coverage	Journals	Research Articles
Astronomy	1979-2008	55	276017
Energy	1978-2008	74	177448
Nanotech	1981-2007	48	81212
Nutrition	1979-2008	69	141274
Oceanography	1979-2008	53	80873

For longitudinal panel analysis, publications were split into 3 major time periods: 1978-1992, 1993-2000, and 2001-2008. The first time period represents the global polarity in the world between the USSR and USA. The second time period was chosen as it is widely recognized that an explosive growth in international scientific publications followed the post-Soviet time period (Georghiou, 1998; Glänzel, 2001), and the last time period is the current state of field.

Whole counting is used for enumerating country publication and measuring international collaboration (Gauffriau 2007). This methodology assigns one credit to all unique countries in a publication. Thus when counting a collaboration, each unique country is assigned one credit with each other unique country in a publication. For example, in a paper with three authors, one author from USA, China and France, the paper increases the total publications for each of these countries by one and increases the collaboration counts of USA-China, USA-France and China-France each by one unit.

[6] Nanotechnology & Nanoscience's coverage is shorter than others as the first journal classified in this subject area started at 1981. Additionally, 2008 data was not available at the time data collection occurred. Data for Nanotech was collected in the spring of 2009 and the rest of the fields during the fall of 2009.

Additionally, it would count as a single publication for both USA, China and France. The advantage of using this method of counting is it focuses on the extent of external collaboration while disregarding internal collaboration. Thus is it focuses on measuring the contribution of a country to a field. The alternative method, fractional or complete counting, attributes credit proportionally so that each individual author is counted separately. Thus for the three-author example, each country would 1/3 credit for the publication. The advantage of this method is that it measures the participation of a country to a field and measures both internal and external collaborations. Since the focus of this research is on external collaboration and on the developing network between countries, whole counting is used.

Methodologies for Evaluation

This research employs four common methods for understanding international collaborations that have been used traditionally in scientometric analysis and introduces a common statistical methodology to the field of Scientometrics that enhances our understanding of international collaboration over a given time period. The four traditional methodologies employed are: collaboration percentages, collaboration intensity, affinity ratios, and citation impact. The new methodology is the odds ratio.

For illustrative purposes, the example of China's output and its interactions with its top collaborators in Energy & Fuels from 2001-2008 will be used. This includes 5,586 documents with Chinese authors, of which 22% (1,188) are internationally coauthored.

Percentage of collaboration

The most common and oldest method for understanding collaboration (Beaver & Rosen 1978) is to measure the number of papers published by a country that also claim an affiliation with at least one other country divided by the total output of that country. This measurement provides a macro indicator showing how internationalized a country's scientific program is.

Figure 5 shows the percentage of collaboration for the G8 countries + China and South Korea in Energy & Fuels across all three time periods under study in this paper. This graph shows a macroscopic view of the international collaboration within the field across several countries. Growth in collaboration among Western countries and Japan and South Korea has grown rapidly at a rate between 500%-700%. In contrast, international collaboration in China grew at a more modest pace increasing from 9% to 20%, or 220%, over the same time period.

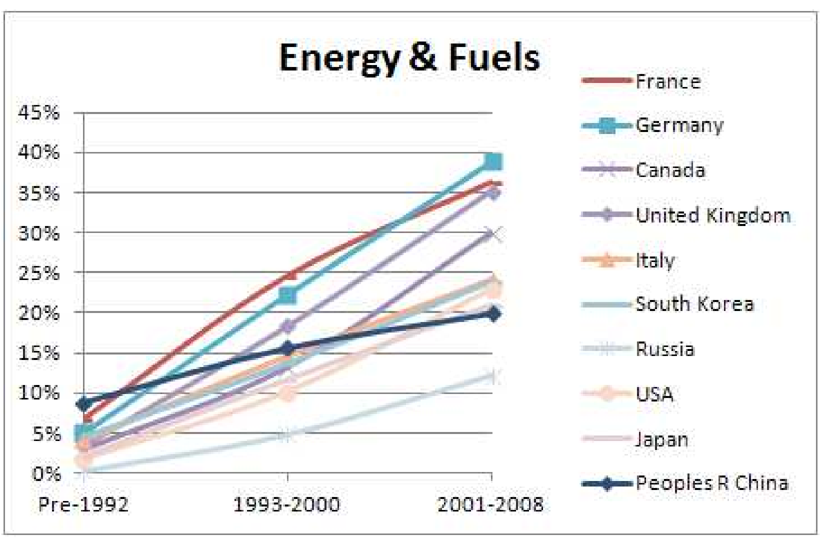


Figure 5 Growth in International Scientific Cooperation as % of total articles in Energy & Fuels

For usage to study collaboration between two countries, Luukkonen, et al (1992) modified this method by limiting the numerator to only publications coauthored between two countries while keeping the denominator the total publication output of the country in question. This indicator can be used to rank the importance of foreign countries contributions by their percentage of collaboration (table 10). When compared with the world ranking for the country (based on article counts), a rough measure can be seen for how important a country is to China in comparison to others. For instance, Singapore, which is ranked 30th in the world in Energy & Fuels, is China's 6th largest collaboration. This signifies that Singapore is more important a partner to China than other top ranked countries such as Germany and France.

Table 10 China's Top 10 Collaborators in Energy & Fuels (2001-2008)

Collab Rank	Country	Count	Percent	World Rank
1	United States	371	6.64%	1
2	Japan	186	3.33%	3
3	Canada	136	2.43%	5
4	United Kingdom	100	1.79%	4
5	Australia	68	1.22%	14
6	Singapore	63	1.13%	30
7	Germany	54	0.97%	6
8	South Korea	51	0.91%	11
9	Taiwan	32	0.57%	15
10	France	29	0.52%	8

Collaboration Intensity

Collaboration intensity (CI), also referred to as Salton's index or the cosine similarity, is used to normalize the amount of bilateral collaboration between two countries and is defined as

Equation 4Collaboration Intensity

$$CI = \frac{C_{xy}}{\sqrt{C_x \times C_y}}$$

Where C_{xy} is the count of papers coauthored together C_x and C_y is the total number papers authored by country X and Y (Glänzel 2001, Wagner and Leydesdorff 2005, Wagner 2008).

CI is interpreted in both absolute and relative terms. In absolute terms CI ranges from 0-1, with a 1 signifying that every paper published by two countries were coauthored together, a very rare occurrence. In practice, many countries have authors who publish without any international collaboration, so CI tends to be low and therefore the interpretation is done in comparing the intensity relative to other CI. The relative interpretation is done across countries, within a field, or across time.

Table 11 Collaboration Intensities for China's Top 10 Collaborators in Energy & Fuels (2001-2008)

Rank	Country	Count	Percent	Intensity
1	United States	371	6.64%	4.55
2	Japan	186	3.33%	3.99
3	Canada	136	2.43%	3.38
4	United Kingdom	100	1.79%	2.32
5	Australia	68	1.22%	2.44
6	Singapore	63	1.13%	4.24
7	Germany	54	0.97%	1.36
8	South Korea	51	0.91%	1.57

9	Taiwan	32	0.57%	1.21
10	France	29	0.52%	0.78

The CI is added to the China example for illustrative purposes (table 11). The main difference between the CI and the percentage is that the CI takes into account the size of the scientific output of both collaborative partners. Thus, the intensity of collaboration with Singapore is on par with the United States. This measure can also differentiate two nations that are close in terms of percentages, such as Germany and South Korea. While they are virtually tied in terms of percentage, South Korea has a CI that is 15% greater. Additionally, Australia has almost a third less collaborations with China than the United Kingdom but maintains a slightly higher CI. Thus Australia enjoys a closer relationship with China than the UK does.

Affinity Ratio

The affinity is an asymmetric indicator (Luukkonen, et al 1992, Glänzel 2000 and Glänzel & Schubert 2001) that measures the importance of another country to a selected country. It is defined as the ratio of the percentage share of joint papers (SJP) in the internationally co-authored papers of the select country (C_{IPx}) and the percentage share of the partner country (SPC) in the total world output (C_W) minus country under study:

Equation 5Affinity

$$Aff_{x \to y} = SJP / SPC$$

Equation 6Share of Joint Papers

$$SJP = C_{xy} / C_{IPx}$$

Equation 7Share of Partner Country

$$SPC = {C_y}/{C_w - C_x}$$

100% affinity would indicate no affinity with a partner country. In other words, the select country co-authors with the partner country as much as the rest of the world does. Values greater than 100% show positive affinities while less than 100% would be a negative affinity. So a 200% affinity would mean the partner country is twice as important to the select country as it is to the rest of the world while a 50% affinity would mean it is half as important.

Table 12 Affinity Ratios for China's Top 10 Collaborators in Energy & Fuels (2001-2008)

Rank	Country	Count	Percent	Intensity	Affinity To	Affinity From
1	United States	371	6.64%	4.55	175.82%	150.90%
2	Japan	186	3.33%	3.99	270.08%	291.41%
3	Canada	136	2.43%	3.38	265.40%	205.68%
4	United Kingdom	100	1.79%	2.32	169.82%	112.52%
5	Australia	68	1.22%	2.44	275.30%	172.98%
6	Singapore	63	1.13%	4.24	897.22%	573.35%
7	Germany	54	0.97%	1.36	107.50%	66.96%
8	South Korea	51	0.91%	1.57	152.50%	142.18%
9	Taiwan	32	0.57%	1.21	143.12%	279.85%
10	France	29	0.52%	0.78	65.42%	40.16%

Affinities are useful for providing signals of an important relationship, or lack thereof between two countries. Expanding again on the China example (table 12), further insights in the collaborative behaviors can be gleaned. For example, Singapore is 9 times more important to China than it is to the rest of the world. Conversely, France is nearly a third less important to China as it is to the rest of the world. The assymetrical nature of this measurement can be informative. For instance, while China has no particular affinity towards Germany, Germany has a negative affinity (67%) towards China. And while China has only a slight affinity toward Taiwan (143%), Taiwan has a very strong affinity (280%) toward China. This lopsided behavior is indicative of a dependency on China for Taiwan in the area of Energy & Fuels.

Citation Impact

Citation impact is a relative measure of the mean citation rate of an entity gauged against some aggregated class of entities. For example, in comparing an institution's mean citation rate versus the average citation rate of journals and/or the subfields they have published in or against some sort of global average (Schubert and Braun 1985, Moed et all 1995, Glänzel 2001). Glänzel (2001) refers to this as the Relative Citation Rate (RCR) and provides a general description as:

Equation 8 Relative Citation Rate

$$RCR = \frac{MOCR}{MECR}$$

Where MOCR is Mean Observed Citation Rate and MECR is the Mean Expected Citation Rate. For the study of the citation impact of international collaboration between countries, the MOCR is the mean citation rate of collaboration between the two countries and the MECR is the mean citation rate of the country's publications that were published without any collaboration (similar to the methodology used in Moed 2005). A RCR of 1 means the collaboration has neutral effect, >1 a positive effect and <1 a negative effect.

Table 13 Citation Impact for China's Top 10 Collaborators in Energy & Fuels (2001-2008)

Rank	Country	Count	Percent	Intensity	Affinity To	Affinity From	Citation Impact
1	United States	371	6.64%	4.55	175.82%	150.90%	1.01
2	Japan	186	3.33%	3.99	270.08%	291.41%	1.12
3	Canada	136	2.43%	3.38	265.40%	205.68%	1.17
4	United Kingdom	100	1.79%	2.32	169.82%	112.52%	0.94
5	Australia	68	1.22%	2.44	275.30%	172.98%	0.86
6	Singapore	63	1.13%	4.24	897.22%	573.35%	1.26
7	Germany	54	0.97%	1.36	107.50%	66.96%	1.43
8	South Korea	51	0.91%	1.57	152.50%	142.18%	1.21
9	Taiwan	32	0.57%	1.21	143.12%	279.85%	1.36
10	France	29	0.52%	0.78	65.42%	40.16%	0.55

The citation impact provides a proxy for determining the impact of the collaboration. A high citation impact could be indicative of a relation in which the country is creating new knowledge that is valued by the community. While a low citation impact could be a sign of a technology transfer or application of existing methods to a localized problem. Expanding upon the Chinese example (see table 13), the citation impact is greatest for Germany. Therefore, even though the relationship with Germany is less intense than others, it is producing work that has a significantly higher citation rate than work produced solely by Chinese scientists. Conversely, the work with France has the lowest citation impact. Given that it is the weakest partner on this list, this is confirmation that France is not a strategic partner for China in Energy & Fuels.

Odds Ratio

The Odds Ratio is an asymmetric indicator that measures the growing or lessening of collaboration between two countries over time. The Odds Ratio is a standard statistical measure of an event occurring in one group to the odds of it occurring in another group. In applying this to studying international scientific collaboration, the event is country *x* collaborating with country *y* over a time period *t*. Thus the country odds ratio (COR) is defined as:

Equation 9Country Odds Ration

$$COR_{x \to y} = \frac{\sum_{t=cl(n/2)}^{n} C_{xy}}{\sum_{t=cl(n/2)}^{n} (C_x - C_{xy})} \Bigg/ \frac{\sum_{t=1}^{fl(n/2)} C_{xy}}{\sum_{t=1}^{fl(n/2)} (C_x - C_{xy})}$$

Take the example of an odds ratio analysis of China’s top 10 collaborators in energy & fuels in 2001-2008 (figure 6). For calculating the odds ratio of China towards USA, first take the sum of all China-USA collaborations in 2005-2008 and divide this by the sum of all Chinese publications minus those that were in collaboration with USA in the same time period; and then divide this by the same ratio but instead summing all papers from 2001-2004. This would provide an odds ratio of 0.87. The interpretation is that China is 13% less likely (1 minus the odds ratio) to collaborate with USA in the latter half of the 2001-2008 than in the earlier half.

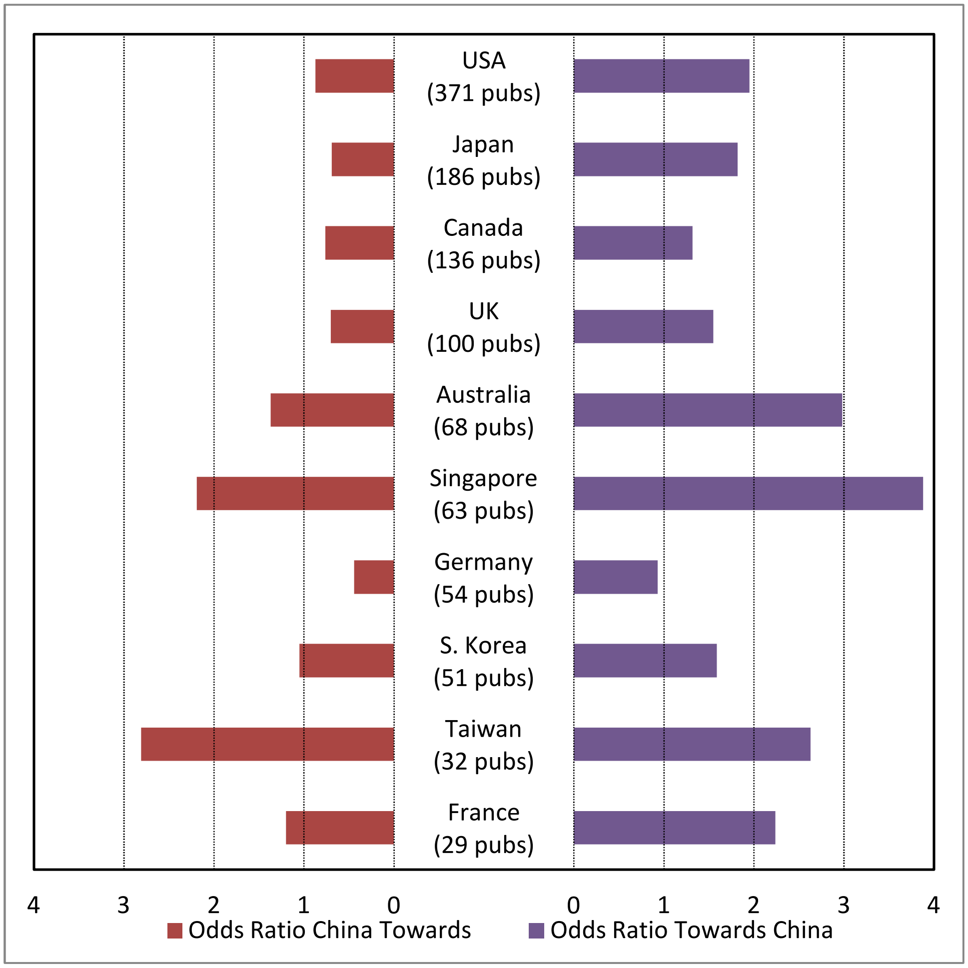


Figure 6 China's Top 10 Collaborators in Energy & Fuels (2001-2008) Odds Ratio Analysis

In figure 6, the bars to the left of the country shows the odds ratio of China towards the given country for collaboration and the bars to right are a given country's odds ratio to collaborate with China. This graph shows that countries with a presence in oceans south of China are an increasing growth area for China (odds ratio > 1) while others are of lessening significance (odds ratio < 1). While at the same time, for all countries save Germany, China is increasingly being sought after as a collaborating partner.

Table 14 Odds Ratio for China's Top 10 Collaborators in Energy & Fuels (2001-2008)

Rank	Country	Count	Percent	Intensity	Affinity To	Affinity From	Citation Impact	Odds Ratio To	Odds Ratio From
1	United States	371	6.6%	4.55	175.8%	150.9%	1.01	0.87	1.95
2	Japan	186	3.3%	3.99	270.1%	291.4%	1.12	0.69	1.82
3	Canada	136	2.4%	3.38	265.4%	205.7%	1.17	0.76	1.32
4	United Kingdom	100	1.8%	2.32	169.8%	112.5%	0.94	0.7	1.55
5	Australia	68	1.2%	2.44	275.3%	173.0%	0.86	1.37	2.98
6	Singapore	63	1.1%	4.24	897.2%	573.4%	1.26	2.19	3.88
7	Germany	54	1.0%	1.36	107.5%	67.0%	1.43	0.44	0.93
8	South Korea	51	0.9%	1.57	152.5%	142.2%	1.21	1.05	1.59
9	Taiwan	32	0.6%	1.21	143.1%	279.9%	1.36	2.81	2.63
10	France	29	0.5%	0.78	65.4%	40.2%	0.55	1.2	2.24

In conclusion, all these methods shine a different light unto the relationship between China and its partners (see table 14). For example, China’s currently maintains a strong affinity to collaborate with Japan but the odds ratio points towards a decline in their relationship as China grows. In contrast, China is a weaker affinity with Taiwan but it is growing much stronger than other relations as it collaborates on projects that are having a large citation impact. However, citation impact alone is not reason enough for growth as the odds ratio for Germany is showing.

Case Studies

Each of the methodologies presented above provides a different type of insight that can be used in conjunction to evaluate a country's scientific relationship with the world. Five case studies that examine a country's collaboration over time are briefly presented to illustrate that point.

The first two case studies, United States in Energy & Fuels and France in Oceanography, provide examples of how international collaboration analysis may be used to evaluate the relationship among scientifically advanced countries and how they relate to the developing world. The next two case studies, Chile in Astronomy & Astrophysics and South Korea in Nanoscience & Nanotechnology, demonstrate how the evaluation of international collaborations sheds light on how a country rises from the periphery of science to the core. The final case study of Nutrition research in Africa demonstrates how to evaluate the relationship between the developed and the developing nations.

Each case study contains at the minimum two pieces of information: 1) details of the general publication trends of the country or countries under study, and 2) an international collaboration analysis of the top collaboration partners. The international collaboration analysis is performed separately for each time period specified in the case study and follows the format outlined in table 15 accompanied by a discussion of the results of the analysis. There is a slight variation of this in the African study in which the rank was dropped as many countries had only a few significant collaboration partners (defined greater than 5 collaborations). For this case study, all African countries were combined into two tables with a column listing each country. See the case study on Nutrition in Africa for more details.

Table 15 International Collaboration Analysis Legend

Rank	Country	Count	Percent	Intensity	Citation Impact	Aff. To	Aff. From	Odds Ratio To	Odds Ratio From
Rank by count	Collab-partner	# of pubs	% of total output	Collab intensity	Citation impact compared to national pubs	Aff. of case study to collab partner	Aff. of collab partner to case study	Odds Ratio of case study to collab partner	Odds Ratio of collab partner to case study

United States in Energy & Fuels

Table 16 US Output in Energy & Fuels

	Pubs	% Collab	Avg Cites
pre-1992	11557	5%	5.4
1992-2000	7153	16%	11.7
2001-2008	11900	22%	7.4

The United States is the global leader for research in Energy & Fuels with over four times the number of publications than its closest rivals in field, China and the UK. The global average for papers with international collaboration in energy and fuels was 14% in 2001-2008 (Zelnio 2012), the lowest among all five fields analyzed. Among the G8 + China & South Korea, the range for collaboration during that time period is between 12% and 38% with the USA being close to the median (see figure 5). Even with this low collaboration percentage, the USA is an ideal candidate for a case study in this field due to the sheer volume of publications (table 16).

Table 17 Top 15 US International Collaborations Analysis in Energy & Fuels (pre-1992)

Rank	Country	Count	Percent	Intensity	Citation Impact	Affinity To	Affinity From	Odds Ratio To	Odds Ratio From
1	Canada	129	1.1%	3.5	0.8	762%	320%	0.8	1.1
2	United Kingdom	80	0.7%	2.4	1.7	610%	220%	1.8	1.2
3	Australia	58	0.5%	2.8	2.5	1078%	321%	1.0	0.7
4	France	40	0.4%	1.7	5.7	607%	201%	1.9	0.8
5	China	25	0.2%	2.1	2.0	1471%	380%	37.4	21.2
6	Japan	23	0.2%	1.1	1.3	446%	336%	2.0	1.1
7	Italy	22	0.2%	1.2	1.4	553%	227%	1.5	1.1
8	Netherlands	18	0.2%	1.4	0.5	929%	191%	1.4	0.9
9	Egypt	17	0.2%	1.7	0.3	1316%	298%	2.2	0.7
9	Israel	17	0.2%	1.4	2.4	877%	298%	1.3	0.8
11	New Zealand	15	0.1%	1.7	1.9	1570%	360%	0.8	0.7
12	Brazil	14	0.1%	1.4	1.5	1107%	304%	0.6	0.5
13	Switzerland	12	0.1%	1.3	3.2	1238%	274%	0.5	0.2
14	Mexico	11	0.1%	1.5	0.2	1666%	386%	0.6	0.4
15	Germany	9	0.1%	0.9	0.8	696%	171%	13.9	0.1
15	Hungary	9	0.1%	1.8	6.0	2786%	456%	13.9	3.8
15	Sweden	9	0.1%	1.0	1.7	854%	293%	0.4	0.2

In the pre-1992 time period international collaborations analysis (table 17), the top three collaborators are all from English speaking countries, with the top being the US's northern neighbor. The collaborations with France are at a lesser intensity than others in the top-5 but the citation impact of the research was the greatest.

The sheer geographic diversity of countries sets the United States apart from other countries. This diversity reflects patterns noted in the US maintaining ties in both Europe and Asia (section 3.2) and its strong presence in Africa (section 4.1). Even here in the early time period analysis, the United States has strong affinities and mostly positive odds ratio towards peripheral non-western countries such as China, Egypt, Israel, New Zealand, Brazil, Mexico and Hungary.

Table 18 Top 15 US International Collaborations Analysis in Energy & Fuels (1992-2000)

Rank	Country	Count	Percent	Intensity	Citation Impact	Affinity To	Affinity From	Odds Ratio To	Odds Ratio From
1	United Kingdom	140	2.0%	3.8	0.9	290%	199%	2.2	1.6
2	Canada	122	1.7%	4.0	1.0	367%	322%	1.5	1.5
3	France	110	1.5%	3.9	1.7	376%	218%	1.6	1.6
4	Germany	102	1.4%	3.6	1.2	348%	207%	2.0	1.5
5	Japan	99	1.4%	3.0	1.1	253%	299%	4.1	2.5
6	Australia	70	1.0%	3.2	1.3	415%	338%	4.5	3.7
7	China	55	0.8%	2.6	1.1	330%	278%	2.5	0.7
8	Russia	45	0.6%	1.7	1.0	170%	292%	2.3	0.9
9	Israel	39	0.6%	3.0	0.9	643%	353%	1.4	1.5
10	Netherlands	38	0.5%	2.2	1.3	359%	193%	0.8	0.6
11	Spain	32	0.5%	1.5	1.9	207%	161%	2.4	2.1
12	Sweden	30	0.4%	1.7	1.0	278%	181%	2.2	1.1
13	Switzerland	29	0.4%	2.4	1.4	542%	266%	1.4	0.9
13	Venezuela	29	0.4%	2.6	0.4	659%	413%	0.5	2.3
15	Mexico	27	0.4%	2.3	0.5	525%	356%	6.7	3.2

In the 1992-2000 analysis (table 18), the UK overtakes Canada as the top collaborator as the US strengthens its ties to Western Europe in general and brings their intensities in-line with those shared by Canada and Australia. The largest jump up in the ranks is Germany whose strong growth in the odds ratio in pre-1992 continues into this time and it leaps forward to 4th place from 15th. The US also dramatically increases its odds ratio with Japan as it brings its collaboration intensity in-line with those of its European partners. Israel also maintained a large intensity with the US during this time.

During this time the US forged strong affinities with Russia, increasing the odds of collaboration. This push would prove unsustainable as Russia fell to 24th rank in 2001-2008 due in part to Russia's lack of engagement in collaboration with the world in this field (see figure 5). Venezuela followed a similar pattern in which the high collaboration intensity formed during this time period dissolved and it fell to rank 26 in the 2000s.

Table 19 Top 15 US International Collaborations Analysis in Energy & Fuels (2001-2008)

Rank	Country	Count	Percent	Intensity	Citation Impact	Affinity To	Affinity From	Odds Ratio To	Odds Ratio From
1	China	371	3.1%	4.6	0.8	151%	176%	2.0	0.9
2	Canada	276	2.3%	4.7	0.9	217%	196%	1.2	0.9
3	United Kingdom	200	1.7%	3.2	1.1	137%	106%	1.2	1.2
4	Germany	198	1.7%	3.4	1.5	159%	115%	1.1	1.1
5	South Korea	179	1.5%	3.8	1.2	216%	234%	1.9	1.3
6	Japan	168	1.4%	2.5	1.3	98%	124%	0.8	0.9
7	France	151	1.3%	2.8	1.2	137%	98%	1.0	0.9
8	Australia	110	0.9%	2.7	1.6	179%	131%	0.7	0.7
9	Turkey	93	0.8%	1.7	0.9	86%	203%	2.0	1.5
10	Italy	81	0.7%	1.9	1.8	121%	129%	1.7	1.2
11	Mexico	70	0.6%	2.4	1.1	215%	232%	0.7	0.9
12	Netherlands	69	0.6%	2.0	1.1	162%	108%	1.1	0.8
13	India	66	0.6%	1.2	2.0	58%	112%	1.1	0.7
14	Spain	62	0.5%	1.3	1.3	69%	67%	0.7	0.5
15	Switzerland	59	0.5%	2.2	1.1	217%	139%	1.0	0.9

2001-2008 (table 19) saw the dramatic rise of Asia as a collaboration partner for the USA. A strong growth in collaborations with China propelled it to the top of the list with an intensity that rivals that of Canada. South Korea also makes a strong debut at the 5th rank with intensity greater than any of those with Europe. Japan keeps a strong relation with the United States though the odds ratio dips slightly below 1.0 for the first time as the affinities for collaboration between these countries levels off. Rounding out Asia growth is the introduction of India to the top 15 as it bypasses Spain (whose less negative odds ratio and affinity with the US helped it drop 3 ranks).

Western European countries remain a strong partner with the US through maintaining a steady relation during this time period. The large growth in odds ratio by Mexico in 1992-2000 pays off with it the two countries maintaining a strong intensity and mutual affinity during the 2000s. The US also grew its collaborations with Turkey significantly during this time, making Turkey its premier collaborator in the Islamic world. Egypt, its premier partner in pre-1992, has steadily declined to 27th rank in 2001-2008, just behind Venezuela.

France in Oceanography

Table 20 France output in Oceanography

	Pubs	% Collab	Avg Cites
pre-1992	361	25%	32.2
1992-2000	1251	45%	30.8
2001-2008	2233	60%	12.4

France is the fifth largest producer of articles in oceanography, borders both the Atlantic Ocean and the Mediterranean Sea, has several colonies in the South Pacific and Indian Ocean and is a leading European Union (EU) nation. This case study of France in the field of Oceanography helps shed insight as to how France has juggled these factors over time.

Table 21 Top 5 France International Collaborations Analysis (pre-1992)

Rank	Country	Count	Percent	Intensity	Citation Impact	Affinity To	Affinity From	Odds Ratio To	Odds Ratio From
1	United States	36	10.0%	2.8	1.3	184%	411%	0.7	1.0
2	United Kingdom	13	3.6%	2.7	1.1	475%	662%	1.1	1.5
3	Canada	9	2.5%	1.7	1.7	265%	306%	2.1	2.1
4	Italy	7	1.9%	3.6	0.9	1536%	1424%	0.3	0.3
5	French Polynesia	5	1.4%	7.9	0.6	10373 %	4769%	4.1	3.0

Pre-1992, France's collaborations were limited to a handful of nations (table 21). The most predominant of these was its relationship with the United States, whose collaborations were greater than the next top 4 collaborators combined. France also maintained a strong intensity with its neighboring countries the UK and Italy and disposition to collaborate with its colony French Polynesia.

Table 22 Top 15 France International Collaborations Analysis (1992-2000)

Rank	Country	Count	Percent	Intensit y	Citatio n Impact	Affinity To	Affinity From	Odds Ratio To	Odds Ratio From

1	United States	128	10.2%	4.5	1.4	85%	153%	1.2	1.6
2	United Kingdom	110	8.8%	7.1	1.3	253%	327%	1.8	1.7
3	Germany	65	5.2%	5.1	1.3	221%	244%	0.7	0.7
4	Spain	47	3.8%	5.6	1.0	363%	422%	1.9	1.5
5	Canada	46	3.7%	3.4	1.4	139%	206%	0.8	1.2
6	Italy	42	3.4%	6.0	1.3	465%	508%	1.4	0.9
7	Netherlands	39	3.1%	4.5	1.9	283%	333%	0.5	0.4
8	Belgium	29	2.3%	6.2	1.3	729%	582%	0.7	0.6
9	Japan	25	2.0%	2.8	1.2	169%	291%	0.3	0.3
10	Australia	24	1.9%	2.0	1.2	95%	181%	1.2	1.0
10	New Caledonia	24	1.9%	10.7	0.9	2639%	1678%	1.0	0.5
12	Norway	21	1.7%	2.7	1.4	190%	230%	1.3	1.3
13	Denmark	18	1.4%	2.4	1.4	178%	173%	2.2	2.1
14	Greece	16	1.3%	4.2	2.0	591%	597%	11.3	4.5
15	French Polynesia	15	1.2%	8.7	1.1	2749%	1749%	0.6	2.3
15	Portugal	15	1.2%	4.4	1.0	709%	616%	10.6	5.3

The 1990s (table 22)saw a shift in France's priorities away from its collaborations with North America to focus on building strong ties to EU nations. While the United States remains France's number 1 partner and an area of growth, France's affinity for collaboration with the US crossed the threshold and went negative as it dipped below 100%. France collaboration intensities with its bordering EU countries grew to surpass the United States. The odds of collaborating with Greece and Portugal have sky rocketed showing a concerted effort by France to include these southern European countries.

France increased its presence in the Pacific during this time period, albeit at a lower rate than in Europe. It maintained its strong affinity to collaborate with French Polynesia but substantially decreased its odds of collaboration. France has instead intensified it collaboration with its other substantial South Pacific colony, New Caledonia. France also maintains a low intensity relationship with Japan and Australia though its Japanese collaborations has a very low odds ratio score which is reflective of how little importance Japan is to France in relation to Japan's general importance to world.

Table 23 Top 25 France International Collaborations Analysis (2001-2008)

Rank	Country	Count	Percent	Intensity	Citation Impact	Affinity To	Affinity From	Odds Ratio To	Odds Ratio From
1	United States	332	14.9%	6.7	1.4	73%	98%	1.3	1.4
2	United Kingdom	278	12.5%	9.8	1.5	185%	202%	1.0	1.3
3	Germany	137	6.1%	5.8	1.7	133%	140%	0.9	1.1
4	Spain	136	6.1%	7.3	1.4	211%	270%	1.0	0.9
5	Netherlands	116	5.2%	7.4	1.3	253%	269%	1.4	1.7
6	Italy	96	4.3%	6.3	1.4	220%	258%	1.1	1.3
7	Belgium	93	4.2%	8.9	1.9	459%	427%	0.8	0.8
8	Canada	88	3.9%	4.0	1.2	95%	111%	1.2	1.4
9	Australia	87	3.9%	3.8	1.5	90%	148%	2.2	2.3
10	Norway	70	3.1%	4.4	1.6	145%	169%	1.6	1.6
11	Japan	54	2.4%	2.7	2.2	70%	116%	0.7	0.8
12	South Africa	46	2.1%	5.2	1.8	312%	362%	1.0	1.2
13	Sweden	45	2.0%	3.6	1.5	155%	173%	0.8	1.0
14	Portugal	44	2.0%	4.1	1.9	201%	259%	0.7	0.5
15	Greece	42	1.9%	5.4	1.3	368%	479%	1.1	1.0
16	Denmark	38	1.7%	2.9	1.6	122%	120%	1.3	1.6
17	Russia	36	1.6%	1.8	0.7	46%	97%	1.9	2.9
18	Monaco	30	1.3%	7.2	1.9	938%	668%	0.6	0.6
19	New Caledonia	30	1.3%	7.8	0.8	1094%	809%	4.4	5.9
20	Chile	26	1.2%	3.0	0.8	181%	210%	1.2	1.1
21	Peru	26	1.2%	7.8	0.6	1251%	909%	7.4	6.0

22	China	23	1.0%	1.3	0.9	39%	71%	1.4	0.9
23	New Zealand	23	1.0%	1.6	1.0	62%	106%	3.2	4.9
24	Switzerland	20	0.9%	3.1	2.0	252%	213%	0.6	0.8
25	Brazil	19	0.9%	2.0	1.2	117%	161%	1.0	0.8

The third time period, 2001-2008, was a turning point for France in that the majority of all publications now were international co-authored (table 20). This follows a trend shared by most EU countries towards greater reliance on collaboration (see Figure 5). The international collaborations analysis (table 23) shows a greater increase in collaboration intensities with all its neighboring EU countries, though there is a significant lag in the intensity with Germany in comparison with other neighbors. The United States continues to retain its top rank in collaborations but affinities between the two countries continue to decline. This decline in affinities may be symbolic that while France cannot ignore its partnership with the top country in oceanography research, it has chosen to continue its focus on bolstering its presence in Europe.

Outside of Europe and North America, France has bolstered its presence in Africa forming intensive collaborations with South Africa and Monaco. It has recently started increasing its odds ratio of collaboration with Russia and the research with Japan, while not as intense as others, has the largest citation impact of all its collaborative partnerships.

The extended analysis into the top 25 partners shows France is actively growing its presence in the South Pacific. It is increasing its relation with Australia both in terms of intensity and in its odds ratio, which is the highest among developed nations. While its colony in French Polynesia has continued to decline to the point it does not even make the top 25 list, France has increased its relations with New Caledonia as well as significantly increasing the odds of its collaboration with neighboring New Zealand. Across the Pacific, France has dramatically grown a relationship with Peru, becoming Peru's most significant foreign partner.

Chile in Astronomy & Astrophysics

Chile provides a unique case study of how a developing country positions itself to be a member of the core countries involved in Astronomy & Astrophysics (A&A) (Zelnio 2012). They have accomplished this by exploiting their geographic advantage: the Atacama Desert. The Atacama is the driest places on Earth with an average rainfall in the Chillean region of just 1mm. This aridity and its elevation has made it one of the most ideal spots for astronomy helping to seal its selection for the European Southern Observatory (ESO) (Duerbeck 2003). ESO has built and currently operates 8 telescopes in Chile (table 24).

Table 24 ESO Telescopes in Chile

Name	Size	Began Operations
ESO 3.6 m telescope	3.57 m	1977
MPG/ESO 2.2 m telescope	2.20 m	1984
New Technology Telescope (NTT)	3.58 m	1989
Very Large Telescope (VLT)	4 x 8.2 m + 4 x 1.8 m	1998
Atacama Pathfinder Experiment (APEX)	12 m	2005
Visible and Infrared Survey Telescope for Astronomy (VISTA)	4.1 m	2009
Atacama Large Millimeter/submillimeter Array (ALMA)	50 x 12 m, and 12 x 7 m + 4 x 12 m	2011

VLT Survey Telescope (VST)	2.6 m	2011

Chile's output of publications has grown exponentially over the past 30 years (Table 25). However, this growth has been driven purely by international collaboration. Domestic production has been flat at ~200 papers per time period. While collaboration rates are generally high in A&A, Chile's percentage has risen at a rate significantly higher than other developed countries in the core (see Figure 5). Its average citation rate is the highest of the top 26 countries that comprises the core of A&A.

Table 25 Chile Output in A&A

	Pubs	% Collab	Avg Cites
pre-1992	534	63%	40.1
1992-2000	1130	81%	53.6
2001-2008	2864	93%	26.7

In the first time period (table 26), Chile was heavily dependent on the United States with a history of collaboration that spans a century (Duerbeck 2003). However, the effects of ESO were already evident with strong collaboration intensities with France and Netherlands and strong growth with the UK and Italy. Its highest citation impacts were with the Netherlands and its primary Southern Hemisphere partner, Australia. In Latin America, Chile maintained strong affinities to Mexico and Brazil.

Table 26 Top 15 Chile International Collaboration Analysis in A&A (pre-1992)

Rank	Country	Count	Percent	Intensity	Citation Impact	Affinity To	Affinity From	Odds Ratio To	Odds Ratio From
1	United States	179	33.5%	5.4	1.8	188%	414%	1.4	1.9
2	France	53	9.9%	4.5	1.6	431%	572%	1.2	1.2
3	Netherlands	38	7.1%	4.3	2.6	557%	573%	0.9	1.1
4	United Kingdom	35	6.6%	2.4	1.6	184%	245%	2.5	3.3
5	Italy	32	6.0%	2.8	0.8	269%	418%	2.8	2.7
6	Canada	29	5.4%	2.8	1.4	316%	494%	0.7	0.8
7	Australia	21	3.9%	2.8	2.4	427%	489%	0.4	0.5
8	Spain	18	3.4%	2.6	1.3	434%	565%	1.9	1.1
9	Sweden	18	3.4%	3.7	0.7	849%	836%	2.9	3.1
10	Germany	16	3.0%	2.5	1.2	454%	546%	2.5	0.0
11	Mexico	12	2.3%	3.0	0.4	831%	1237%	10.9	7.5
12	Brazil	11	2.1%	2.4	1.5	573%	976%	3.0	1.6
12	Japan	11	2.1%	1.2	0.8	142%	316%	1.6	1.2
12	Switzerland	11	2.1%	2.2	0.8	486%	593%	1.6	2.0
15	Belgium	10	1.9%	2.3	0.7	622%	733%	2.7	3.4

The 1990s saw Chile become even closer to the United States even as collaboration with Europe grew (table 27). The effects of ESO are even more pronounced as collaboration intensity grew and the odds of collaboration were higher with all Europe with the exception of Italy. The largest change was the growth of collaboration with Germany, where it grew from rank 10 in pre-1992 to the second most important partner in 1992-2000.

Table 27 Top 15 Chile International Collaboration Analysis in A&A (1992-2000)

Rank	Country	Count	Percent	Intensity	Citation Impact	Affinity To	Affinity From	Odds Ratio To	Odds Ratio From
1	United States	482	42.7%	8.2	1.7	156%	195%	1.7	1.3
2	Germany	207	18.3%	6.6	1.7	234%	256%	2.1	1.3
3	France	167	14.8%	6.2	1.0	258%	295%	1.8	1.3

4	Italy	129	11.4%	4.8	0.9	202%	257%	1.0	0.8
5	United Kingdom	123	10.9%	3.9	2.1	140%	165%	1.5	1.0
6	Spain	89	7.9%	4.7	2.3	271%	312%	1.5	0.9
7	Netherlands	86	7.6%	4.7	0.9	287%	294%	1.7	1.4
8	Australia	78	6.9%	4.9	4.6	344%	394%	2.7	1.9
8	Canada	78	6.9%	4.1	1.4	243%	302%	0.9	0.9
10	Mexico	54	4.8%	4.7	0.9	459%	539%	1.9	0.9
11	Belgium	47	4.2%	5.2	0.9	638%	815%	1.2	0.8
12	Brazil	45	4.0%	3.5	0.9	307%	451%	1.3	0.7
12	Sweden	45	4.0%	4.3	1.0	453%	487%	1.6	1.1
14	Russia	40	3.5%	1.8	0.6	89%	131%	1.8	0.8
15	Austria	27	2.4%	3.5	1.0	505%	577%	3.8	1.8
15	Poland	27	2.4%	2.3	1.2	227%	293%	1.0	0.6

The collaborations with Australia in the 1990s continue to be of significance as it has the highest citation impact. The two countries maintain a high affinity for collaboration that grew at a rate higher than most. Chile is actively growing its collaboration with Russia, however such a low intensity and a negative citation impact signals that this collaborative relationship is perhaps more about establishing ties than conducting new science.

Chile continues to maintain a strong mutual affinity with Mexico and Brazil in the 1990s but the odds ratio from these countries have leveled off and the payoff in terms of citation impact is minimal.

Table 28 Top 15 Chile International Collaboration Analysis in A&A (2001-2008)

Rank	Country	Count	Percent	Intensity	Citation Impact	Affinity To	Affinity From	Odds Ratio To	Odds Ratio From
1	United States	1427	49.8%	12.3	3.6	124%	122%	0.8	1.1

2	Germany	825	28.8%	12.1	3.1	206%	213%	1.0	1.1
3	France	585	20.4%	10.1	3.2	203%	219%	1.0	1.0
4	United Kingdom	563	19.7%	8.1	3.6	137%	150%	1.4	1.5
5	Italy	472	16.5%	8.1	2.2	165%	195%	1.2	1.3
6	Spain	349	12.2%	7.9	3.1	210%	234%	1.0	1.1
7	Canada	248	8.7%	6.1	4.4	174%	207%	1.4	1.3
8	Australia	225	7.9%	6.6	3.0	227%	265%	1.2	1.4
9	Netherlands	221	7.7%	6.0	3.3	188%	195%	1.3	1.5
10	Argentina	157	5.5%	8.0	1.3	481%	647%	2.1	2.9
11	Poland	151	5.3%	5.4	4.1	228%	294%	0.9	1.1
12	Brazil	146	5.1%	5.1	1.6	209%	359%	1.2	1.4
13	Belgium	144	5.0%	6.3	1.9	325%	392%	1.3	1.2
14	Sweden	141	4.9%	5.9	4.9	289%	331%	0.9	1.1
15	Russia	139	4.9%	2.8	1.9	64%	98%	1.2	1.5

The 2000s (table 28) continued to see the role of the United States grow in importance as collaborations with the US now accounted for 50% of all of Chile’s output, though its odds of collaborating with the US in relation to the rest of the world has become negative. The same pattern of increased collaboration intensity but a leveling off of the odds ratio is seen with Chile’s European partners. This means that while Chile is increasing its collaboration with the world, it is doing so uniformly now with countries in the Western Hemisphere. The only country in the top 15 that Chile is growing with is a new entrant to the list, its geographic neighbor Argentina. As Argentina is a peripheral country (Zelnio 2012), this increase in collaboration is most likely political rather than scientifically motivated.

The 2000s also mark a turning point for Chile in that all collaborations now have a positive citation impact. This, coupled with the very high collaboration rate shown in table 25, demonstrates that Chile has become completely reliant on collaboration for advancing its scientific base in A&A.

South Korea in Nanoscience & Nanotechnology

Table 29 South Korean output in Nanoscience & Nanotechnology

	Pubs	% Collab	Avg Cites
Pre-1992	7	0%	4.7
1992-2000	604	19%	8.7
2001-2008	3092	24%	4.0

South Korea in Nanoscience & Nanotechnology (N&N) provides a case study on how a peripheral country with almost no publications in a field twenty years ago rises to over 3,000 publications in the most recent time period (Table 29). In comparison to the world, South Korea rose from being ranked 24th in terms of publication output in 1991 to 5th in 2007 (table 30). The way this was accomplished was through a build-up of internal capacity and strategic partnerships with key leaders in the field.

Table 30 South Korea Publication Rank over Time in Nanotechnology & Nanoscience

Rank	country	# Documents	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
1.	United States	19591																	
2.	Japan	7807																	
3.	China	7240																	
4.	Germany	6328																	
5.	France	4293																	
6.	United Kingdom	3759																	
7.	South Korea	3700																	
8.	Taiwan	2322																	
9.	India	2157																	
10.	Italy	2132																	
11.	Spain	1703																	
12.	Canada	1641																	
13.	Russia	1539																	
14.	Singapore	1326																	
15.	Switzerland	1282																	
16.	Netherlands	1172																	
17.	Belgium	1080																	
18.	Australia	1015																	
19.	Sweden	955																	
20.	Brazil	823																	
21.	Poland	772																	
22.	Austria	670																	
23.	Israel	585																	
24.	Czech Republic	552																	
25.	Greece	495																	

In table 30, there are two predominant strategies shown in N&N collaboration. The first strategy, typically followed by European countries, shows a steady growth in collaboration ending in 40%-50% of all publications in 2001-2007 being internationally coauthored. The second model is the US & Asian model in which growth in collaboration is limited with international collaboration in 2001-2007 accounting for 20-% 25% of total output. This model, followed by South Korea, relies on a much stronger domestic base in which collaborations are targeted.

Table 31 Top 5 South Korea International Collaborations Analysis (1992-2000)

Rank	Country	Count	Percent	Intensity	Citation Impact	Affinity To	Affinity From	Odds Ratio To	Odds Ratio From
1	United States	67	11.1%	3.9	1.2	263%	222%	1.6	8.6
2	Japan	18	3.0%	1.6	0.9	177%	180%	0.2	0.9
3	China	12	2.0%	1.7	1.9	287%	208%	1.5	2.6
4	Russia	8	1.3%	1.7	0.5	423%	160%	1.0	1.5

5	United Kingdom	6	1.0%	0.8	0.4	124%	75%	0.8	2.9

South Korea's collaboration in 1992-2000 (table 31) was concentrated on achieving a strong relationship with the United States and, to a much lesser extent, its neighboring countries and exclusion of any significant collaboration with Europe.

The strategy of collaboration with Japan has mixed signals. While Japan is the second ranked country, the intensity and affinity are lower than other neighbors and the odds of collaboration with Japan in relation to the rest of the world are very low. These odds are reflective of the fact that the Japanese during this time period is the second most prolific authors in N&N and account for a much larger share of the worldwide output yet only account for 3% of South Korean output. A possible reason for this is that both Japan and South Korea are pursuing a strategy of low cooperation and strong domestic output.

Table 32 Top 10 South Korea International Collaborations Analysis (2001-2007)

Rank	Country	Count	Percent	Intensity	Citation Impact	Affinity To	Affinity From	Odds Ratio To	Odds Ratio From
1	United States	338	10.9%	5.1	1.2	154%	113%	1.2	1.4
2	Japan	160	5.2%	3.8	1.0	183%	167%	0.9	1.5
3	China	81	2.6%	1.8	1.3	82%	85%	1.0	0.7
4	Germany	49	1.6%	1.3	0.9	68%	34%	1.2	1.8
5	United Kingdom	44	1.4%	1.5	0.9	103%	58%	0.5	0.6
6	India	32	1.0%	1.3	0.9	112%	123%	4.0	3.1
7	Russia	23	0.7%	1.2	0.8	128%	53%	1.9	4.1
8	Taiwan	22	0.7%	0.9	0.4	71%	129%	0.7	0.5
9	Australia	19	0.6%	1.2	0.9	150%	84%	1.1	1.2
10	France	14	0.5%	0.5	0.5	29%	14%	0.8	1.1

In 2001-2007 time period (table 32), South Korea maintained its close ties with the United States. During this time period they greatly expanded their collaboration with Japan, doubling their intensity. A factor in this growth was Japan's increased odds of collaborating with South Korea. South Korea also expanded its collaboration efforts to target Germany, the top European N&N research country, and to a lesser extend the UK and France. It also expanded its collaboration with regional partners, maintaining steady growth with China and Russia and open up new collaborations with India, Taiwan and Australia. From a citation impact, South Korean gained the most advantage from its collaborations with USA and China and the least from Taiwan and France. Both Taiwan and France and leading researchers in N&N (table 30), so the low citation impact and low intensity are probably signals that the collaborations were more about establishing relations than a push to conduct new ground-breaking research.

Nutrition in Africa from 2001 to 2008

The case study of nutrition research in Africa was chosen as an example of the relationship between the developed and developing world as nutrition science is considered "a foundation for development" by the United Nations and is considered a critical problem identified in the Millennium Development Goals (United Nations Standing Committee on Nutrition, 2009). Every country in Africa is considered in the periphery of nutrition scientific research (Zelnio 2012).

The nutrition data for this case study was limited to only publications that contained at least one author from an African nation in the years 2001-2008. This resulted in a total of 1,466 publications. The distribution of these articles across African countries is shown in figure 8. For the collaboration analysis, only countries with at least 20 articles, of which 5 are international collaborations with another country, were considered. Only Sudan had more than 20 articles and failed to meet this criterion with only 4 collaborations with Germany.

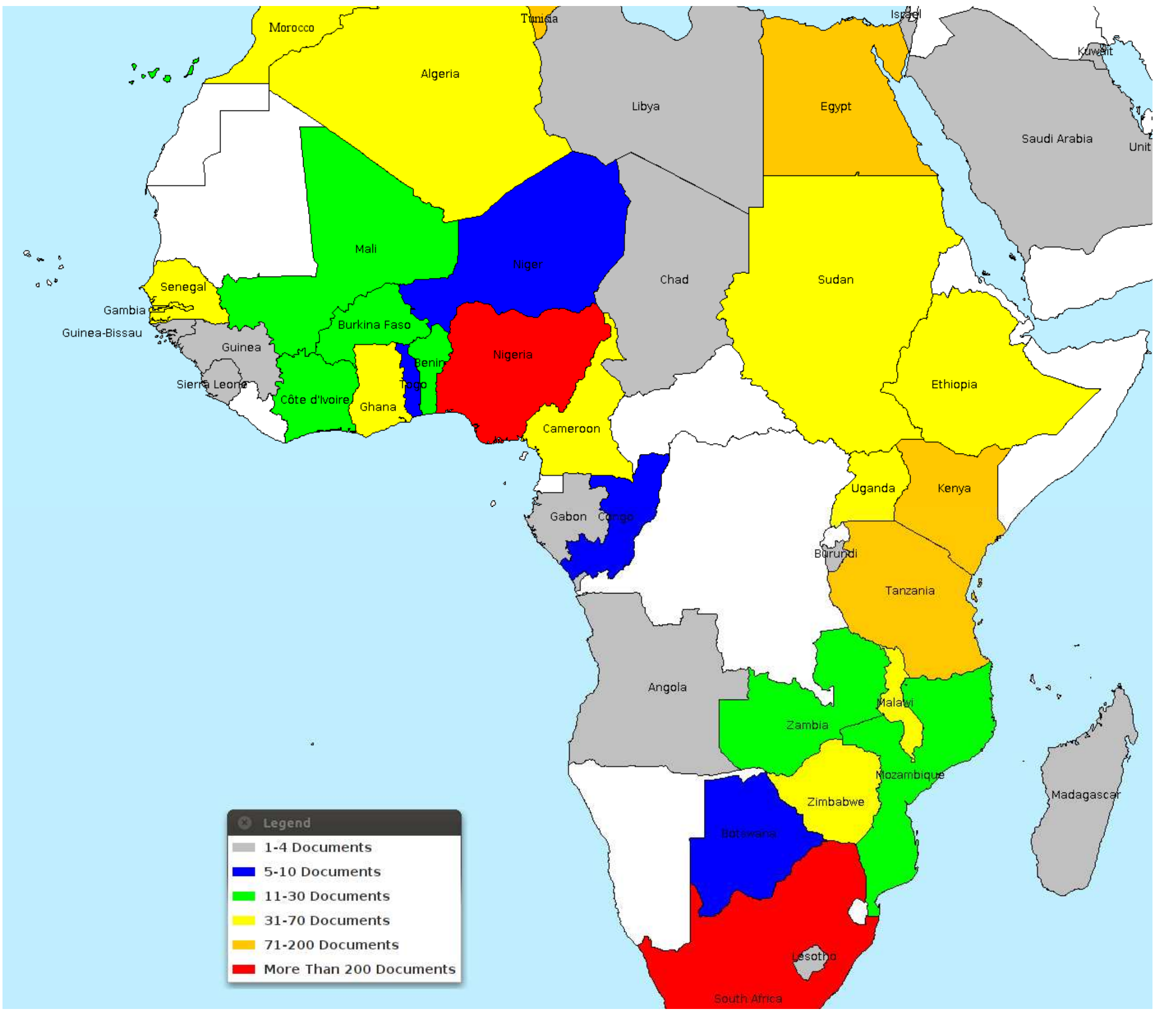


Figure 7 Geographical Distribution of Nutrition Research in Africa (2001-2008)

A co-country network analysis of this data (figure 9) reveals a continent divided along languages relating to its colonial history and its linguistic legacy. The co-country network depicts connections between countries that have co-authored together at least 5 times, with the thickness of the line being representative of the number of collaborations and the size of the node being the total number of publications within this data set. Figure 9 shows that French-speaking Africa is dominated by France, while the rest of Africa is dominated by the United States and the United Kingdom, with little to no overlap between these two divisions. French Africa is exclusive in its collaboration with French-speaking countries (France, Belgium and Switzerland) with the most notable exceptions being Morocco, which has strong ties to its European neighbor Spain, and Cameroon, who's Northwest Province and Southwest Province are English-speaking.

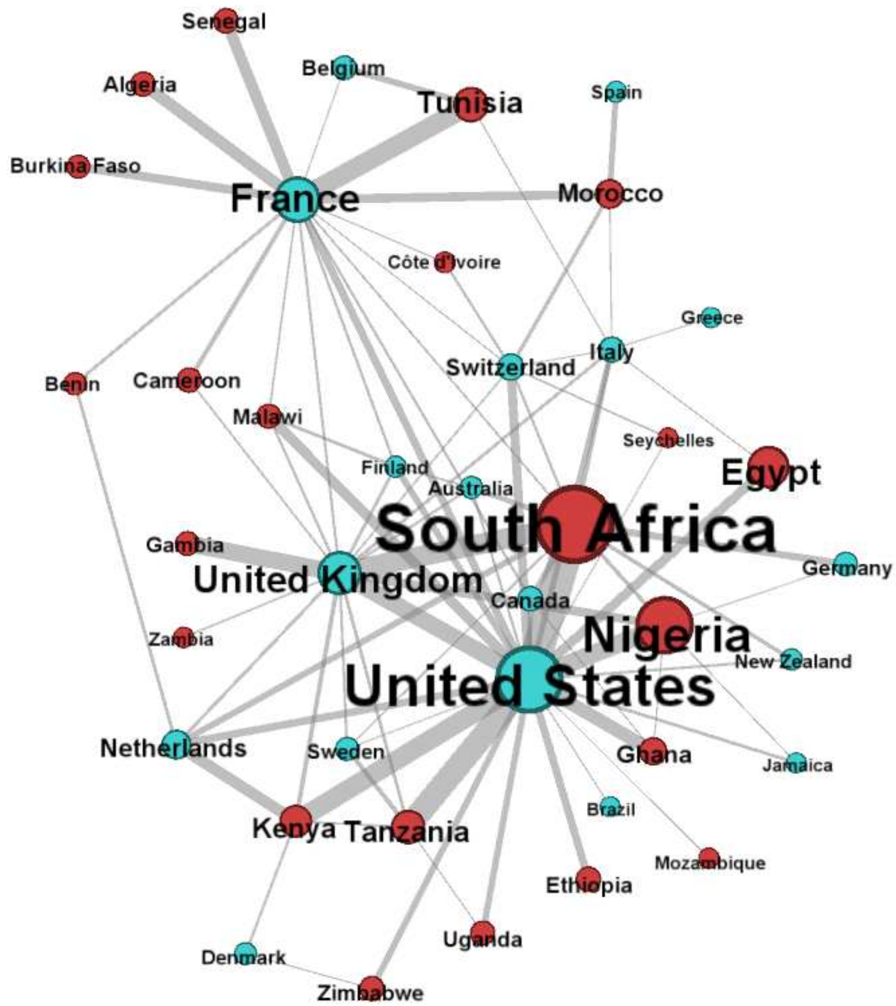


Figure 8 Network of African Collaboration in Nutrition (2001-2008)
(red = Africa, blue = non-Africa)

The international collaboration analysis is split into two separate tables to reflect the division between English (table 33) and French (table 34) speaking Africa.

Table 33 International Collaboration Analysis of English Speaking Africa

African Country	Country	Count	Percent	Intensity	Citation Impact	Affinity To	Affinity From	Odds Ratio To	Odds Ratio From
S. Africa	United	54	15.7%	2.1	2.1	126%	170%	2.3	3.0

(344 - 42%)	States								
S. Africa (344 - 42%)	UK	36	10.5%	2.5	1.7	265%	292%	0.5	0.6
S. Africa (344 - 42%)	Germany	14	4.1%	1.4	1.2	207%	227%	3.4	3.9
S. Africa (344 - 42%)	Netherlands	13	3.8%	1.5	1.4	263%	233%	1.0	1.2
S. Africa (344 - 42%)	Canada	12	3.5%	1.2	4.0	184%	205%	0.6	0.8
S. Africa (344 - 42%)	Australia	11	3.2%	1.3	2.9	245%	296%	2.6	2.7
S. Africa (344 - 42%)	Italy	10	2.9%	1.0	2.6	164%	207%	2.3	2.1
S. Africa (344 - 42%)	New Zealand	9	2.6%	2.1	3.5	773%	785%	2.0	2.1
S. Africa (344 - 42%)	Nigeria	9	2.6%	3.2	1.0	1835%	2388%	3.8	8.7
S. Africa (344 - 42%)	Switzerland	7	2.0%	1.1	3.8	266%	191%	0.5	0.5
S. Africa (344 - 42%)	France	6	1.7%	0.6	1.3	86%	87%	2.4	2.5
S. Africa (344 - 42%)	Sweden	6	1.7%	0.8	4.0	182%	160%	1.2	1.3
Nigeria (225 - 32%)	United States	32	14.2%	1.5	2.1	148%	154%	2.0	1.3
Nigeria (225 - 32%)	UK	17	7.6%	1.4	1.1	249%	211%	1.3	0.6
Nigeria (225 - 32%)	S. Africa	9	4.0%	3.2	1.3	2388%	1835%	8.7	3.8
Nigeria (225 - 32%)	Germany	5	2.2%	0.6	1.5	147%	124%	2.0	1.1

Nigeria (225 - 32%)	Ghana	5	2.2%	4.6	1.8	8775%	4183%	2.0	0.8
Nigeria (225 - 32%)	Jamaica	5	2.2%	4.1	4.6	6913%	2710%	0.5	0.5
Egypt (133 - 34%)	United States	18	13.5%	1.1	1.0	133%	147%	0.3	0.3
Egypt (133 - 34%)	Italy	5	3.8%	0.8	0.9	262%	267%	1.4	0.9
Tanzania (88 - 85%)	United States	41	46.6%	3.1	3.7	182%	505%	1.6	1.7
Tanzania (88 - 85%)	Sweden	10	11.4%	2.7	2.4	581%	1041%	0.7	0.7
Tanzania (88 - 85%)	UK	8	9.1%	1.1	3.3	113%	254%	0.5	0.5
Tanzania (88 - 85%)	Kenya	6	6.8%	7.2	0.5	6753%	7015%	0.8	0.9
Tanzania (88 - 85%)	Uganda	5	5.7%	9.0	0.5	12541 %	14400 %	9.3	8.5
Kenya (78 - 82%)	United States	34	43.6%	2.7	3.1	177%	473%	1.3	1.2
Kenya (78 - 82%)	Netherland s	21	26.9%	5.0	2.3	954%	1658%	0.8	0.7
Kenya (78 - 82%)	UK	10	12.8%	1.4	2.1	165%	358%	3.6	2.4
Kenya (78 - 82%)	Denmark	7	9.0%	2.3	1.9	612%	1040%	1.2	1.0
Kenya (78 - 82%)	Tanzania	6	7.7%	7.2	0.5	7015%	6753%	0.9	0.8
Ghana (52 - 67%)	United States	24	46.2%	2.4	1.6	229%	501%	3.6	3.5
Ghana (52 - 67%)	Canada	5	9.6%	1.3	2.4	314%	565%	1.9	2.5
Ghana (52 - 67%)	Nigeria	5	9.6%	4.6	1.7	4183%	8775%	0.8	2.0
Ethiopia (37 - 76%)	United States	15	40.5%	1.8	2.3	179%	440%	1.1	1.9
Uganda (35-74%)	United States	13	37.1%	1.6	3.3	167%	403%	0.3	0.9
Uganda (35-74%)	Tanzania	5	14.3%	9.0	0.5	14400 %	12541 %	8.5	9.3
Zimbabwe (35 - 74%)	United States	13	37.1%	1.6	1.4	167%	403%	2.2	1.5

Zimbabwe (35 - 74%)	Denmark	5	14.3%	2.5	2.0	1078%	1655%	0.1	0.1
Malawi (33 - 97%)	United States	19	57.6%	2.4	0.6	198%	624%	0.9	2.0
Malawi (33 - 97%)	Finland	8	24.2%	4.6	1.0	1818%	3845%	1.3	2.0
Malawi (33 - 97%)	UK	7	21.2%	1.5	0.5	231%	593%	0.3	0.6
Malawi (33 - 97%)	France	6	18.2%	1.9	1.2	386%	907%	0.8	1.2
Gambia (32 - 97%)	UK	30	93.8%	6.7	9.6	1003%	2687%	0.4	0.9

In this analysis, next to the African country's name in parenthesis are the total number of publications and the percentage of publications that are internationally co-authored. In the English-speaking Africa, only South Africa, Nigeria and Egypt have substantial stand-alone scientific base in nutrition research, the rest have at least 2/3 of the output internationally coauthored. While typically international collaborations are a positive indicator, in this case there are no inherent scientific base which may take advantage of such collaborations.

The intra-Africa collaboration is limited to Nigeria-South Africa, Nigeria-Ghana, Tanzania-Kenya and Tanzania-Uganda. These collaborations are highly intensive, ranging from 3.2 to 9.0, and are growing (odds ratio >= 2.0) or relative stable (odds ration = 1 ± 0.2). The collaborations that involved Nigeria generally had a positive citation impact while those involving Tanzania had a negative impact. This may be reflective of Nigeria's scientific output being 2.5 times larger than Tanzania.

Not surprisingly, the USA and UK dominate as the preferred collaborative partner throughout the region, usually by an order of magnitude to other countries. However, in all instances except in Kenya, the odds ratio of the UK collaborating with an African country is less than 1, meaning the UK is now less likely to collaborate with African nations. In contrast, the odds of the United States collaborating are rising throughout much of English-speaking Africa, thus filling the void left by the UK lessening its interest in the continent. The exception for this is Kenya, where the odds ratio is relatively flat for the US but is significantly high for the UK.

The rest of the developing world targets collaboration with just 1 or 2 select African countries. Germany collaborates with a relatively low intensity and citation impact with South Africa and Nigeria. The Netherlands has a very strong intensity and mutual affinity to collaborate with Kenya which also has a strong citation impact. It also has a relatively weaker collaboration with South Africa. The only incursions of France into English Africa are a very weak presence in South Africa and a surprisingly strong affinity to collaborate in Malawi. Several Scandinavian countries have unusual ties to Africa such as Sweden's strong ties to Tanzania, Finland has a strong affinity to collaborate with Malawi and Denmark's short but strong interactions with Zimbabwe in the early part of the 2000s that was not sustained in the second half of the time period (odds ratio of 0.1). A possible explanation of these small but significant forays of these countries into Africa may be indicative of relationship between a professor and their student that lasted for a few years after the student graduated. A more detailed analysis would be required to confirm this hypothesis.

The one significant outlier in this analysis is the third largest scientific producer in Africa, Egypt. While the United States is its major collaborator, its affinity is positive but weak and the odds of these two countries continuing their relationship is very low. Its ties with its other collaborator, Italy, are growing from Egypt's perspective but the collaboration intensity is low and non-sustainable. Additionally, the citation impact of the research done with both countries is low.

Table 34 International Collaboration Analysis of French Speaking Africa

African Country	Country	Count	Percent	Intensity	Citation Impact	Affinity To	Affinity From	Odds Ratio To	Odds Ratio From
Tunisia (94 - 62%)	France	34	36.2%	6.2	1.9	1206%	1805%	0.4	2.5
Tunisia (94 - 62%)	Belgium	13	13.8%	4.6	1.3	1714%	2103%	1.5	5.3
Tunisia (94 - 62%)	Italy	5	5.3%	1.0	2.1	203%	378%	1.9	6.7
Morocco (61 - 72%)	France	17	27.9%	3.9	1.4	796%	1391%	1.1	0.8
Morocco (61 - 72%)	Spain	12	19.7%	2.8	2.3	587%	1757%	0.7	0.4
Morocco (61 - 72%)	Switzerland	9	14.8%	3.3	4.3	1118%	1382%	0.8	0.6
Morocco (61 - 72%)	Italy	5	8.2%	1.2	0.7	268%	583%	1.4	0.9
Senegal (37 - 78%)	France	21	56.8%	6.1	∞	1492%	2832%	0.2	1.1
Cameroon (36 - (67%)	France	10	27.8%	3.0	0.7	858%	1386%	0.4	1.5
Cameroon (36 - (67%)	UK	6	16.7%	1.3	3.5	264%	466%	0.2	0.7
Algeria (35 - 77%)	France	22	62.9%	6.6	5.8	1678%	3137%	0.4	1.2
Burkina Faso (28- 100%)	France	17	60.7%	5.7	∞	1400%	3551%	0.8	1.8
Benin (24 - 92%)	France	8	33.3%	2.9	∞	966%	2315%	5.3	0.8
Benin (24 -	Netherlands	8	33.3%	3.4	∞	1207%	2467%	0.2	0.1

92%)

French speaking Africa is highly reliant on France for collaboration with no intra-African collaborations and no presence by the United States. None of the countries have a strong national scientific base. The UK's collaboration with the partially English speaking Cameroons is of low intensity and declining. In Tunisia, Belgium and Italy are signaling strong growth which has resulted in Tunisia shifting its own collaboration from France (as seen in its low odds ratio) to these other countries. By contrast, Morocco has strong intensities with Spain and Switzerland but the odds of their relationships growing are negative. The one outlier in this analysis is Benin, which has strong collaboration intensity with the Netherlands that rivals its relationship with France. However, the odds ratio signals that this relationship was only during the early part of the time period and was not sustained to the latter half.

Note that the infinite citation impacts (∞) are reflective of the fact that the countries in question have no citations to papers written with an international coauthor.

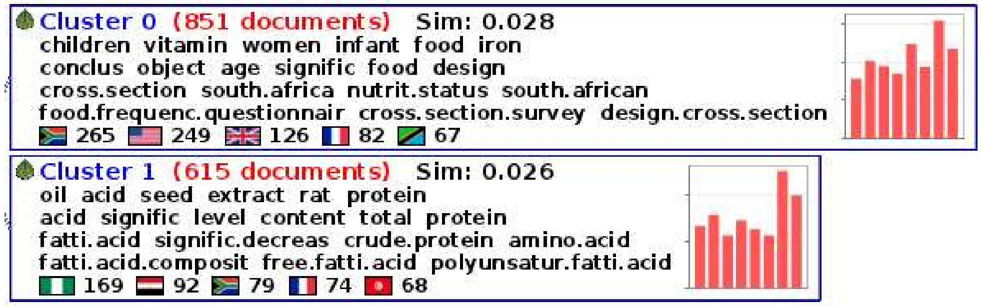


Figure 9 Semantic Clusters of Nutrition Research in Africa

Finally, to gain further insight into the nature of collaboration in Africa, the articles were semantically clustered into two subtopics using the Naval Surface Warfare Center's in-house software package. Figure 9 shows the two clusters along with their key single-, double- and triple word phrases and the top 5 country publication (identified by their flag) counts. Cluster 0 appears to be primarily concerned with the human health aspects of nutrition while cluster 1 is concerned with the nutritional properties of native crops and animals in Africa. The top five in health are South Africa, United States, United Kingdom, France, and Tanzania. The top five in food sources are Nigeria, Egypt, South Africa, France and Tunisia. This reveals that the USA and UK are primarily interesting in collaborating with Africa on researching the nutritional health of Africans while France is evenly split between the health aspects and understanding local food sources.

Summary of Case Studies and Concluding Remarks

The case studies provide diverse examples on the use of international collaboration analysis for the evaluation of national scientific systems. The African case study shows a continent cleanly split by France and the United States. The Chilean case study shows a developing country entering the core by exploiting its geographic advantage to forge close ties to Europe and the United States. The South Korean shows a different path from the periphery to the core by building a strong relationship with its neighbors and the United States. The French case study shows a shift that goes from the strengthening of European Union ties in the 1990s to strengthening ties to the South Pacific in the 2000s. Finally, the case study of the United States shows a strategy of maintaining strong ties to a wide range of countries while strengthening ties to strategic partners like China and Turkey but also weakening ties to Russia and Egypt.

The example of Africa highlights how this may be used by NGOs whose goal is assistance to the developing world, while the cases of Chile and South Korea show differently paths in which a developing country can join the ranks of developed nations. The examples of the United States and France highlight how the analysis provides insights into shifting patterns of collaborations. These examples can thus be used by policy makers to evaluate whether these patterns correspond with national or international priorities.

However, there are limitations to this methodology though that needs be considered. The case studies presented here point towards a need for deeper understanding that the data just simply cannot provide. Why the sudden drop in collaboration between Russia and the increase in collaboration with Turkey and China in the US? Is the shift of France towards the South Pacific indicative of some policy change or rather a reflection of France turning to this area to exploit its access to natural resources? How much has the US's and France's humanitarian aid to Africa resulted in increased scientific ties with those nations or is it all just due to shared language and/or colonial heritages? How much lag is there between the enactments of policies aimed to increase collaboration and actual collaborative articles? This type of analysis cannot answer these questions alone but should accompany qualitative assessments. Going forward, the next step in this research is the need to perform deeper analysis on these case studies to see if there is a link to policy or other social phenomena that can validate the methodologies laid out in this research.

Bibliography

Beaver, D. deB., Rosen R., "Studies in scientific collaboration. Pt 2. Scientific co-authorship, research productivity and visibility in the French scientific elite, 1799-1830", Scientometrics V. 1 (1979) 133-149

Duerbeck, HW. "National and international astronomical activities in Chile 1849-2002" in Interplay of Periodic, Cyclic and Stochastic Variability in Selected Areas of the H-R Diagram. Edited by C. Sterken, ASP Conf. Ser. 292. San Francisco: Astronomical Society of the Pacific, 2003, p. 3-18.

Gauffiau M., Larsen PO, Maye I., Roulin-Perriard A., Von Ins M., "Publication, cooperation and productivity measures in scientific research", Scientometrics, Vol. 73, No. 2 (2007) 175–214

Glänzel W., "National characteristics in international scientific co-authorship relations", Scientometrics, Vol. 51, No. 1 (2001) 69–115

Glänzel W., "Science in Scandinavia: A bibliometric approach", Scientometrics, 48 (2000) 121–150. (Correction: Scientometrics, 49 (2000) 357.

Glänzel W., Schubert A., “Double effort = double impact? A critical view at international co-authorship in chemistry”, Scientometrics, 50 (2001) 199–214.

Glänzel W., Debackere K., Meyer M., “’Triad’ or ‘tetrad’? On global changes in a dynamic world”, Scientometrics, Vol 74, No. 1 (2008) 71-88

Hill, C.T. (2007). “The Post-Scientific Society.” Issues in Science & Technology, 24 (1).

King DA, “The Scientific Impact of Nation”, Nature Vold 430 (2004) 311-316

Leydesdorff L, Wagner CS, “Is the United States losing ground in science? A global perspective on the world science system.” Scientometrics, Vol 78, No. 1 (2009) 23-36

Luukkonen T., Persson O., Sivertsen G., “Understanding Patterns of International Scientific Collaboration”, Science, Technology, & Human Values, Vol. 17, No. 1 (1992) 101-126

Moed H. F., De Bruin R. E. and Van Leeuwen T. N. “New bibliometric tools for the assessment of national research performance: database description, overview of indicators and first applications.” Scientometrics, 33 (1995) 381–422

Moed H. F. (2005). “Does International Collaboration Pay?” in Citation Analysis in Research Evaluation. Springer: The Netherlands

Schott T., “Global Webs of Knowledge: Education, Science, and Technology”, American Behavioral Scientist, Vol 44, No. 10 (2001) 1740-1751

Schubert A. and Braun T,.”Relative indicators and relational charts for comparative-assessment of publication output and citation impact”. Scientometrics, 9 (1986) 281–291.

United Nations Standing Committee on Nutrition. “Why Nutrition is important.” *United Nations Standing Committee on Nutrition*. 2009. Web. 18 Apr. 2012

Wagner, CS (2008). The New Invisible College: Science for Development. Brookings Institution Press: DC

Wagner CS, Leydesdorff L, “Network structure, self-organization, and the growth of international collaboration in science”, Research Policy 34 (10), p.1608-1618 2005

Zelnio, R., “Identifying the global core-periphery structure of science.” Scientometrics, Vol.91, No. 2 (2012) 601-615

CHAPTER FOUR: QUANTITATIVE EXPLORATORY HORIZON SCANNING OF EMERGING COUNTRY CAPABILITIES IN SCIENCE & TECHNOLOGY

Abstract

Horizon scanning is a technique growing in use by policy makers as an input into future planning for a variety of subjects. One area in which its use is growing is planning future science and technology investments. The horizon scan is used in this context to gather intelligence on the scientific, technological and innovative activities of nations. This paper introduces a technique that can be used for this purpose. The Quantitative Exploratory Horizon Scan is designed to detect the early emergence of new critical scientific and technical (S&T) topics that may radically alter a country's capabilities and give it an advantage over other nation. This paper introduces this methodology and then applies it to a case study of the United States from 2005-2010 using a scientometric data source.

Introduction

In today's competitive world, intelligence on the scientific, technological and innovative activities of nations is vital to policy makers. From a national security perspective, policy makers need to be aware of possibly disruptive capabilities that may give one nation a competitive advantage over another. From a scientific perspective, policy makers need to ensure that they remain on the scientific frontier. And from a humanitarian perspective, policy makers need to be aware of possible transformative research that may radically alter the health and well-being of a populace. Therefore, identifying and detecting the early emergence of new country capabilities in science and technology is increasingly becoming priority for policy-makers (Georghiou and Harper 2011, Ahlqvist et. al. 2012, Konnola et al. 2012).

There are two different communities of scholars focusing on methodologies to identify and detect emerging trends in science and technology: forecasters and information scientists. In recent years, the forecasting community developed the horizon scanning methodology for emergence detection (van Rij 2010). The horizon scan is focused on detecting weak signals and extrapolating their probability of emerging. By contrast, information scientists focus exclusively on data mining technologies that can detect clusters of information that exhibit emerging characteristics over time.

Rather than competing approaches, these two approaches can be complementary within a limited scope in which there is sufficient data to be mined to house sufficient weak signals over a long enough time period for gathering enough statistics for an extrapolation to be made. The scope of this research is to detect the early emergence of new critical scientific and technical (S&T) topics that may radically alter a country's capabilities and give it an advantage over other nations.

Horizon Scanning

Horizon scanning is an increasingly popular method employed by government agencies in the policy planning process (Schultz 2006, van Rij 2010). Countries and NGOs around the world are investing in research groups performing horizon scans on emergence of new advancements in medicine, security, transportation, environmental science, and energy (Sutherland and Woodroof 2009, Butter et al 2009, van Rij 2010, Botterhuis 2010, Amanatidou et al 2012, Konnola et. al. 2012).

In 2002, The UK's Department for Environment Food and Rural Affairs defined horizon scanning as "The systematic examination of potential threats, opportunities and likely future developments which are at the margins of current thinking and planning."[7] Its aim is the detection of 'weak signals' in order to provide early warning about important future changes (Schultz 2006) and to improve the robustness of policies by identifying gaps in knowledge (van Rij 2010). The focus on weak signals is a core concept of the horizon scan that dates back to early work by Ansoff (1975) and is defined as: "warnings (external or internal), events and developments that are still too incomplete to permit an accurate estimation of their impact and/or to determine their complete responses" (Hiltunen 2008).

Weak signals come from a variety of sources. They can come from data collection (such as web searches), data mining (such as publications or patents), conferences, expert/stakeholder workshops, and surveys (most frequently Delphi Surveys) (Schultz 2006, Amanatidou et al 2012). These weak signals can then be extrapolated using statistical methodologies as well as through qualitative impact assessment techniques to identify emerging issues (Schultz 2006). Emerging issues that have a low probability but a high impact are identified as "wild cards". These wild cards are events that would radically alter the landscape (Amanatidou et al 2012).

[7] http://horizonscanning.defra.gov.uk/ accessed 5/22/2012. See also van Rij 2010.

Approaches to horizon scan are differentiated between exploratory and issue-centered scanning (Amanatidou et al 2012). Exploratory scanning examines a wide range of data to identify emerging issues. Issue-centered scanning is a more focused and uses a core description of an issue to identify data for analysis of emerging issues. These approaches can be used in conjunction, with an initial exploratory scan being used to identify issues for a more in-depth analysis.

Horizon scanning is distinguished from other foresight methodologies in both scope and function (van Rij 2010). The scope tends to be wider and its focus on identifying emerging trends rather than forecasting as it gives no particular time frame. Its systemic approach allows for it to have a repetitive character that allows for validation of issues identified in previous scans in addition to the identification of new emerging issues. Thus its function is highly supportive of forecasting and policy activities as it may be used to select and scope issues for more in-depth analysis and agenda setting.

Quantifying Emergence through Data Mining

Glänzel and Thijs (2012) state that detecting emergence is one of the greatest challenges currently facing scientometrics, a subdivision of information sciences that focuses on scientific literature. Emergence is a term often used but rarely defined in information sciences. Instead it is frequently defined by its characteristics. The two characteristics commonly examined are newness and rapid growth (Cozzens et al 2010, Glänzel and Thijs 2012). Newness can include a new coherent structure that appears over time or it could also refer to a sudden shift in topical focus of an existing cluster. Rapid growth can refer to a sudden growth in either publication or citation activity of a topic over a short period of time (2-5 years).

There are three common data-mining techniques used for clustering: text-based, author-based and citation-based. The most common text-based clustering algorithm used is co-word clustering which uses co-occurrence of words or phrases to associate documents. Author-based techniques use co-authorship analysis to cluster data based on the underlying social structure. Citation-based techniques can use either co-citation or bibliographic coupling. Co-citation uses the co-occurrence of citations of an older document within a newer document to associated documents that are often cited together. Bibliographic coupling however looks the co-occurrence of citations between documents to associate documents that cite similar works.

To detect emergence, the data is cut into time-slices that vary in sizes of 1-5 years in length. These slices are then compared to find clusters that are either new or rapidly growing. To trace clusters over time, a measure of similarity between one slice to the next is chosen, usually based on the clustering methodology. For example, in co-authorship, the similarity in authors is used, in text-based it is terms and in citation-based methods it is cited documents.

While this discussion of techniques is quite generic, they have been applied broadly in the study of emerging topics in recent time period, either by themselves or in conjunction with each: co-word (Yang et al 2012), co-authorship (Bettencourt 2008), co-citation (Chen 2006, Upham and Small 2010), co-citation + co-word (Chen and Guan 2011), and bibliographic coupling + co-word (Glänzel and Thijs 2012).

Scoping

The scope of this research is to detect the early emergence of new critical scientific and technical (S&T) topics that may radically alter a country's capabilities and give it an advantage over other nations. The approach undertaken was to develop a quantitative exploratory horizon scanning (QEHS) methodology focused on text-data mining technologies that are largely autonomous and utilize domain-agnostic algorithms to reduce multiple signals to inform and prioritize a country's emerging S&T topics.

This approach has several benefits. First, the focus on text-data mining technologies allows for the exploration of very large and diverse data sets in a largely autonomous way, allowing for fast turn-around and frequent scanning. The use of domain-agnostic algorithms for signal reduction allows for broad coverage of the S&T spectrum and the focus on prioritizing allows policy-makers and analysts to focus their limited attention span on a subset of data. These prioritizations can easily feed into larger activities such as identifying issue-centered horizon scans, scenario planning, and focused expert surveys.

There are disadvantages to this approach. First is the limitation of the data to be mined. Signals can come from a variety of sources, with text being a limited subset. There is also a time lag associated with text in that by the time most data sources become available, they are already several years old. This is true in large data sets used in S&T analysis such as bibliographic and patent data. Recent advancements in text-data mining of more real-time data such as twitter and micro-blogs are addressing this issue but this research does not employ those techniques. Because of these limitations, the focus on the use text-data mining tends towards topics that are emerging in the near-term.

The narrow scoping of this project minimizes these disadvantages to some extent allowing the benefits to outweigh the disadvantages. However, these limitations need to be kept in mind when the results of this method are applied towards policy-making.

Methodology

QEHS is taking a fundamentally different approach to the question of emergence than the methodology use by the information sciences. Rather than identifying topics that that are emerging over multiple time-slices, the QEHS tries to identify topics that may *potentially* emerge within a single time-slice. Thus the QEHS methodology thus can be thought of detecting "pre-emergence".

Similar to methodologies for detecting emergence(Upham and Small 2010, Glänzel and Thijs 2012), the QHES begins with a basic assessment of the current fundamental research base using data mining techniques to cluster the output into subtopics. However, due to its scope, the assessment is done primarily at the country level (though this method can also be applied at the institution or disciplinary level).

Where it differs is that QHES (figure 10) uses statistical analysis on the clusters to detect emerging areas of growth. The various statistics gathered can then be fused together to produced a rank ordering of the clusters to prioritize a country's emerging S&T topics. This following subsection outlines how this is accomplished.

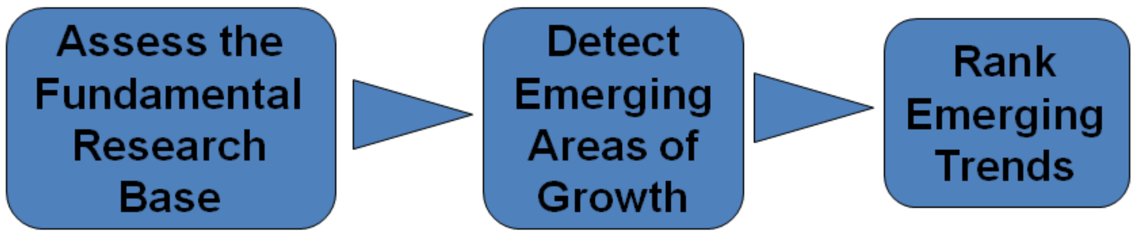


Figure 10 The Horizon Scan Method

Assessing the Fundamental Research Base

There are three factors to be considered for the assessment: time window, data sources, and clustering methods.

The selection of a time window is a non-trivial task. The time window has to be large enough that enough of the research is of the entity under study is captured and trends over time can be calculated. However, there is an upper limit of around 10 years in which the literature has aged to the point terminology is either ubiquitous or has morphed over time. Previous studies of emergence have looked at clustering time slices of 3 years (Glänzel and Thijs 2012) and 6 years (Upham and Small 2010) without any discussion as to what is the ideal time span. For better statistical analysis, the QEHS has erred on the upper part of the range and chosen 6 years as the time span.

The most common data source used to assess a country's research base is one of the large databases of peer-reviewed bibliometric articles such as Thomson Reuters Web of Science of Elsevier's Scopus. Inclusion of one of these databases is adequate to perform an initial horizon scan but for more accurate results, additional information should be included such as patents, dissertations, twitter and academic blogs. The design of the methodology is built so that it can accept data from multiple streams.

There are various techniques which are used for clustering sub-topics. Most common techniques cluster using co-occurrences of words/phrases, citations, authors or some hybrid approach of these co-occurrences (Bettencourt 2008, Shibata et al 2009, Boyack & Klavans 2010, Upham and Small 2010, Glänzel and Thijs 2012). However, the clustering methodology is driven largely by the desire to accept data from multiple streams. Not all of these data sources have citation information and in many cases, authors cannot be properly disambiguated. Therefore, the QEHS follows the approach of Kostoff et. al. (2007) and uses co-word analysis to cluster subtopics.

The software package CLUTO (Karypis 2006) is used to cluster all data sources. This technique uses a hierarchical phrase frequency cluster algorithm that makes the TF-IDF (term frequency-inverse document frequency) vector space model (while omitting stop words) to measure the similarity between documents to assign them to n-clusters. Under this model, the user decides the total number of clusters to partition. For purposes of QEHS, it was decided through trial and error that a mean frequency of between 200 to 400 documents per cluster is ideal. Therefore, the choice the choice for n-clusters is usually determined by dividing the total number of documents by a value by 300 and rounding up to the nearest 1000s. So for instance, in a QEHS of 700,000 documents, the n-clusters would be set to 3,000.

An advantage of using the TF-IDF is that it identifies the top discriminating terms which differentiate this cluster from other clusters. These terms can be used to automatically label each cluster in a meaningful way (Shibata 2009).

Detect Emerging Areas of growth

The QEHS scope is to find early emerging S&T topics that may give a country an advantage. The focus thus is not to detect clusters of technologies that might have already emerged, but to identify those that are likely to emerge. Thus, we are looking for S&T

topics that have a coherent structure (i.e. can be differentiate as its own cluster); a weak signal (low size); and are rapidly growing. Thus statistics on the size of the clusters and their compound annual growth rate are gathered. The compound annual growth rate (CAGR) is calculated based on a linear fit of the cluster over the life of the index:

Equation 10Compoud Annual Growth Rate

$$CAGR(t_0, t_n) = \left(\frac{C(t_n)}{C(t_0)}\right)^{\frac{1}{t_n - t_0}} - 1$$

where $C(t_0)$ and $C(t_n)$ are the count of articles in the linear fit model of the clusters at time t_0 and t_n. The linear fit model is used to avoid the divide by 0 error that occurs if a cluster was not present at t_0.

As a proxy for determining if the topic will give an advantage to the country, is the purity statistic (Zhou and Karypis 2004, Liu et. al. 2005). Purity is concept from information theory for measuring the correlation between an entity and its environment. The higher the purity, the less interaction an entity has with its environment. It is defined as:

Equation 11Purity

$$Purity = \frac{1}{n} \sum_i p_i$$

where n is the total number of documents in the cluster that have the entity and p_i is the fractional count of the entity in each of the n documents (for example, if a document had 2 German authors, 1 Austrian and 1 French, then the p_i for Germany would be 0.5).

Thus purity when applied to country co-authorship, a country with 100% purity within any given cluster means that they are completely developing the science within their borders. A country with a high purity score in an S&T field is thus developing an internal expertise. This internal expertise may give a technical advantage to a country by allowing it an early lead in the development and access to emerging technologies that may lead to disruption. The exact level of advantage is dependent on the nature of the emerging phenomena. The advantage is low in areas that are easily reproduced or copied. The advantage is high in areas that require specific expertise, equipment or facility. Therefore the purity statistic is only measuring a necessary condition for determining advantage. Qualitative assessments by subject experts will need to be used when interpreting the results of the QEHS to gauge just how much country's purity score will give it an advantage should the field emerge.

Rank Emerging Trends

To produce a ranking, each statistic must be normalized to a unit vector prior to being fused into a single score for ranking. For normalization of any statistic (STAT), the following algorithm is used:

Equation 12Statistic Normalization

$$STAT_N = \frac{(STAT - STAT_{min})}{(STAT_{max} - STAT_{min})}$$

Then, the statistics can be fused together by computing the geometric distance of each statistic from its ideal value. For growth and purity, the ideal value is the maximum value of a unit vector, 1. However, for frequency the weak values are of interest so the minimum value, 0, is ideal. Thus the fusion algorithm is defined as:

Equation 13Fusion Score

$$FScore = \sqrt{{F_N}^2 + (G_N - 1)^2 + (P_N - 1)^2}$$

As the fusion algorithm is defined as the distance from the ideal, the lower the fscore, the closer it is to the ideal score. Clusters are then ranked from the lowest fusion score to the highest.

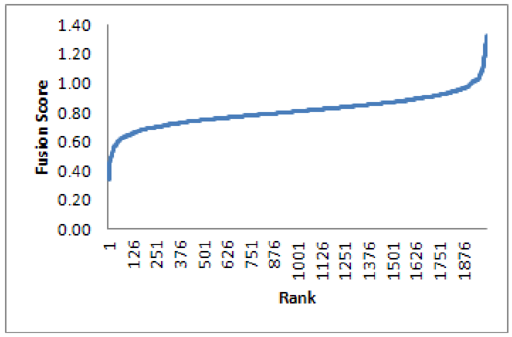
Case Study – United States 2005-2010

A case study of the United States can demonstrate the utility of the QEHS. For this case study, all records published by the United States from 2005 to 2010 in Thomson Reuter's Web of Science were used. The output of the US rose from 470,483 records in 2005 to 495,687 in 2010 for a total of 2,947,016 records. The data was then clustered into 10,000 clusters and table 35 shows the statistics on the cluster solution.

Table 35 US Cluster Statistics

	Freq.	Growth	Purity
Max	5569	228%	100%
Min	20	-56%	43%
Mean	294	2%	88%
Median	201	1%	89%
St. Dev.	316	11%	7%

The combined statistics signals that the distributions of the frequency and growth are highly skewed towards lower values while purity is skewed towards higher values. These heavy tailed distributions produce a heavy tailed fusion score distribution (Figure 11 top). The outliers that make up the left tail (bottom chart in Figure 11) are the weak signals that are of most interest to QEHS.



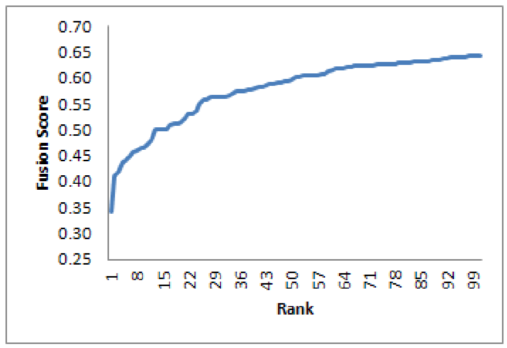


Figure 11 Distribution of Rank vs. Fusion scores.
Top chart is all fusion scores and bottom chart is top 100 scores.

Table 36 shows the top 25 clusters by their fusion score that identify pre-emerging S&T topics as of 2010. Some of these topics are directly related to recent Nobel Prizes. Two of the topics are related to Nobel Prizes in Physics: optical Frequency combs (rank 1) won the 2005 prize and graphene (tied at rank 19) had a major breakthrough in 2004 and won the prize in 2010. Pluripotent research (rank 7 and tied for rank 19), which is at the forefront of embryonic stem cell research, won the 2012 Nobel Prize in Medicine. HPLC mass spectrometry (rank 18) is a new bioanalysis technique that utilizes the electrospray ionization technique, which won the Nobel Prize in Chemistry in 2002. While typically Nobel Prizes are lagging indicators of emergence, the clusters identified by the QEHS show that US researchers are concentrating on maintaining their advantage in these topics.

Table 36 Top 25 Fused Clusters

Rank	Top 5 td-idf phrases	# paper	CAGR	Purity	Fusion
1	comb, frequenc.comb, optic, optic.frequenc.comb, optic.frequenc, waveform	109	228%	95%	0.09
2	fragranc, branch.chain, chain.satur, chain, branch, fragranc.ingredi	41	210%	88%	0.22
2	prostatectomi, robot, robot.assist, robot.assist.radic, assist.radic, assist.radic.prostatectomi	138	176%	93%	0.22
4	medi, synthesi, analog, carbocycl, librari, deriv	52	151%	98%	0.27
5	bodi.radiotherapi, stereotact.bodi.radiotherapi, stereotact.bodi, stereotact, radiotherapi, sbrt	79	133%	96%	0.34
6	patient.center.medic, medic.home, center.medic, center.medic.home, patient.center, medic	57	113%	98%	0.41
7	inor, ligand, complex, complex.support, pyrazolyl, reactiv	233	105%	96%	0.44
7	stem.cell, stem, pluripot, pluripot.stem, pluripot.stem.cell, cell	100	118%	88%	0.44
9	pmse, polym, nanoparticl, polym.surfac, polym.nanoparticl, biodegrad	119	103%	94%	0.45
10	h1n1, influenza, virus, pandem, swine, h1n1.virus	222	122%	84%	0.47
11	epigenet, epigenet.regul, regul, epigenet.mechan, gene, modif	172	102%	88%	0.49
12	crisi, financi.crisi, financi, global.financi, global, global.financi.crisi	151	96%	89%	0.50
13	memori.polym, shape.memori, shape.memori.polym, polym, shape, memori	52	90%	92%	0.51
14	inor, chemistri, synthesi, character, catalyst, oxid	1611	106%	96%	0.52
15	nps, nanoparticl, np, nanoparticl.nps, nm, ag	80	95%	86%	0.53
15	social.media, media, social, public.relat, user, public	59	78%	97%	0.53
17	th17, th17.cell, cell, il-17, th1, cd4	263	93%	86%	0.54
18	anyl, spectrometri, hplc, spectroscopi, chromatographi, separ	271	75%	95%	0.55
19	social.network, social, social.network.site,	141	71%	97%	0.56

	network.site, network, onlin				
19	pluripot, pluripot.stem, pluripot.stem.cell, stem.cell, stem, cell	326	87%	86%	0.56
19	graphen, layer, sheet, graphen.oxid, graphit, graphen.sheet	263	86%	86%	0.56
19	person.medicin, medicin, person, genom, promis, era	80	70%	96%	0.56
19	kidney.injuri, acut.kidney, acut.kidney.injuri, kidney, injuri, acut	277	75%	91%	0.56
19	msc, cell, stem.cell, mesenchym, mesenchym.stem.cell, mesenchym.stem	110	89%	84%	0.56
25	health.care.reform, care.reform, reform, health.care, health, care	325	68%	97%	0.57

Biomedical clusters dominate the top fused clusters, accounting for 12 out of the top 25. These include three stem cell clusters (2 on pluripotent and one on mesenchym – tied at rank 19) mentioned previously, two genetics clusters (th17 – rank 17, and epigenetic – rank 11), four medical research clusters (robotic assisted prostatectomy – tied rank 2, stereotactic body radiotherapy – rank 5, h1n1 influenza – rank 10, and acute kidney injuries – tied at rank 19), and three health care clusters (patient-centered medical homes – rank 6, personalized medicine – tied at rank 19, and health care reform – rank 25).

Semantic clustering is not without its drawbacks. The clusters starting with “medi” (tied at rank 2) and “inor” (tied at rank 7 and rank 14) are byproducts of naming conventions for titles employed by Abstracts of Papers of the American Chemical Society. In both cases, the first term in each cluster were used in all-caps in the title to identify each article. However, this is an exception, not a rule in utilizing this cluster technique and these results can be easily identified during verification. For all three of these, there were similar clusters of much lower rank that these documents would most likely be clustered into without those conventions applied.

Of the remaining 7 clusters, one deals with the toxicity of branched chain chemicals in fragrances (rank 2), 3 deal with polymeric and/or nano- materials (PMSE – ranked 9, shape-memory polymers – rank 13, and silver nanoparticles – tied at rank 15), and the other three deal with social issues (global financial crisis – rank 12, impact of social media – tied rank 15, and online social networks – tied rank 19).

Taken as a whole, this QEHS shows that the United States has several promising fields in biomedical and nano-material science emerging. The three social issues and the three health care emerging clusters are also good barometers on what are the major social issues facing the country under study. With the exception of the global financial crisis, all these clusters have a 97% or higher purity. Coupled with the high growth rate, this is indicative of internal problems facing the nation.

Empirical Validation

The intent of the QEHS is to identify emerging areas of S&T topics that can give it an advantage in the global marketplace. The case study of the United States has identified several promising candidates using the QEHS approach. An analysis of two S&T topics identified in the top 25, optical frequency combs (OFC) and T Helper-17 (commonly referred to as TH17) cells, help illustrate how to validate the results of the case study.

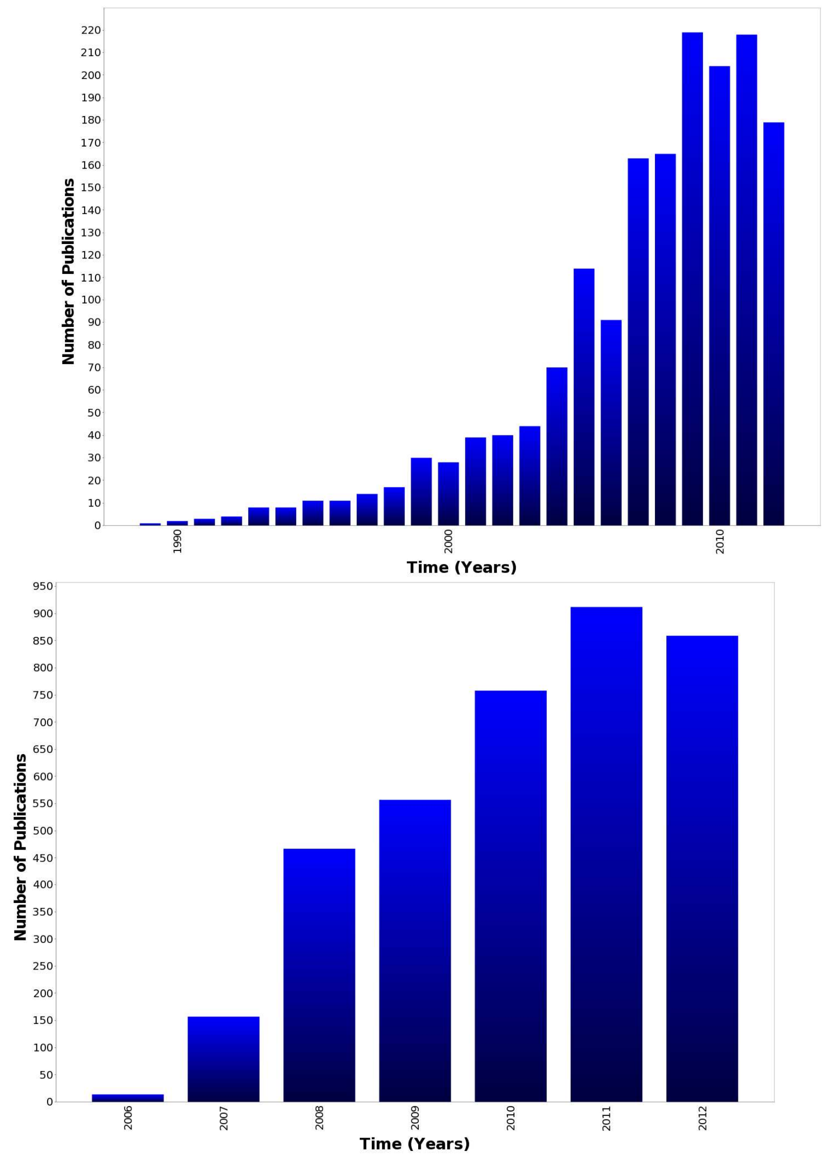


Figure 12 Number of Optical Frequency Comb (top) and TH17 (bottom) Publications (1989-2012)

Web of Science was queried with the phrase “optical frequency comb” on December 12, 2012 and 1,683 results were retrieved. “TH17” was queried on the same day and produced 3,725 results. Growth in OFC was slow until just before the time the Nobel Prize was awarded in 2005 in which growth became exponential (figure 12). The production of articles is slightly more than 200 publications per year from 2009-2011 (2012 data is incomplete), demonstrating that this is still a young, emerging field of study that is growing rapidly. In contrast, the discovery of TH17 cells in 2006 sparked an immediate surge in interest and publications quickly climbed to 912 in 2011, thus showing a field rapidly emerging.

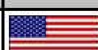
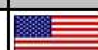
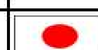
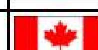
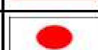
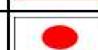
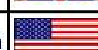
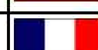
Table 37 Top 10 Countries in OFC and TH17

Optical Frequency Combs					T Helper-17			
Rank	Country	Total Records	H-Index		Rank	Country	Total Records	H-Index
1	United States	557	50		1	United States	1529	94
2	Japan	256	24		2	China	452	28
3	Germany	220	39		3	Japan	412	48
4	China	118	11		4	United Kingdom	334	45
5	United Kingdom	102	20		5	Germany	312	45
6	France	86	21		6	Italy	199	36
7	South Korea	82	12		7	France	171	22
8	Italy	66	15		8	Netherlands	155	25
9	Canada	62	13		9	Canada	135	22
10	Taiwan	57	9		10	Switzerland	115	31

Two bibliometric indicators are used to determine if the US has an advantage over other countries in these two fields: the total records published and the H-index. The H-index is a proxy for determining quality based on peer recognition through a ratio of citations to publications that has been used for evaluating countries (Csajbok et al, 2007) and institutions (Van Raan 2006).

In the case of OFC, the United States has a substantial lead in publication volume, with 1/3 of all papers on OFC coming from its institutions (Table 37). This publication rate is more than double its closest competitors, Japan and Germany. However, Germany's high H-index makes it a strong competitor within this field. In looking at the top institutions (table 38), we see that for most countries, the research is happening at government labs (University of Colorado high rate is due primarily to the fact that 79% of its publications are co-authored with NIST). While NIST has a clear lead within the field, if Germany's two government labs are combined, they provide more evidence of Germany as a strong competitor within this field. Japan also has a strong publication record in OFC but its H-index is considerably lower than Germany's. There are several US universities in the top 10, though their H-index scores are considerably lower than both domestic and foreign government labs. This may be indicative of a high cost barrier to entry in this field to performing cutting edge research which often leads to higher citation publications.

Table 38 Top 10 Institutions in OFC and TH17

Optical Frequency Combs						T Helper-17				
Rank	Institution	Flag	Total Records	H-Index		Rank	Institution	Flag	Total Records	H-Index
1	natl inst stand tech		181	40		1	harvard univ		152	37
2	max planck inst quantum opt		95	29		2	univ tokyo		74	22
3	univ colorado		86	31		3	mcgill univ		68	17
4	purdue univ		67	13		4	niaid		65	26
5	natl inst adv ind sci tech		51	11		5	osaka univ		65	21
6	univ cent florida		43	8		6	kings coll london		60	16
7	korea res inst stand sci		37	7		7	univ pittsburgh		49	16
8	phys tech bundesanstalt		37	16		8	yale univ		49	21
9	natl phys lab		35	12		9	inserm		48	13
10	univ tokyo		34	9		10	nyu		46	20

The advantage the United States has in TH-17 research is much more substantial. It produces 41% of the articles and nearly three times more than its closest competitors. Unlike OFC, the research is driven primarily through research universities though the National Institute of Allergy and Infectious Diseases (NAIAD) is the second leading institute in the United States. Half of the top ten institutes are from the US and the H-index of these institutes are larger than those of similar sizes. While China is the second largest producer of TH-17 research articles, none of its institutions make the top-ten list. The closest competitor is Japan, which has 2 universities in the top 10 with substantially higher h-indices than other non-US institutions of similar size. Thus the advantage that the United States enjoys is quite substantial in this field of research.

Conclusions

Quantitative Exploratory Horizon Scanning has the potential to be a useful tool in to detect the early emergence of new critical S&T topics that may radically alter a country's capabilities and give it an advantage over other nations. The case study of the United States, along with validation, showed that QEHS can identify weak signals that identify emergence in several important S&T topics as well as identify several emerging social issues. Its ability to prioritize the issues allows for a reduction in the data in which to focus resources on identifying S&T topics that may be disruptive and/or may be 'wild cards'. These results of this analysis can be then fed into a much large process for policy planning and foresight activities.

There are many challenges and opportunities ahead. How to identify and integrate data sources that are more recent is one of the key challenges. What other signals in the data can be fused into the analysis to produce more accurate results? How sensitive is the analysis to the time window? There are other potential valuable techniques that may be used to refine the QEHS methodology, such as other clustering techniques for assessing the fundamental research base and other statistics that can be calculated for detecting emergence. Most notably would be the inclusion of citation data in the analysis and other measurements of time trends beyond growth such as burstiness or currency measurements. This paper is meant to be a foundation from which this work can continue.

Bibliography

Ahlqvist T, Valovirta V, and Loikkanen T (2012). "Innovation policy roadmapping as a systemic instrument for forward-looking policy design", *Science and Public Policy* 39: 178–190

Amanatidou, E., Butter, M., Carabias, V., Konnola, T., Leis, M., Saritas, O., Schaper-Rinkel, P. and van Rij, V. (2012) 'On concepts and methods in horizon scanning: Lessons from initiating policy dialogues on emerging issues', *Science and Public Policy*, 39: 208–21.

Ansoff, H. I. (1975) "Managing strategic surprise by response to weak signals", California Management Review, 18: 21–33.

Bettencourt, L.M.A. , Kaiser, D.I. , Kaur, J., Castillo-Ch´vez, C. and Wojick, D.E. (2008). "Population modeling of the emergence and development of scientific fields", Scientometrics, 75: 495–518.

Boyack, K. W., & Klavans, R. (2010). "Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately?" Journal of the American Society for Information Science and Technology, 61(12), 2389–2404.

Botterhuis, L., van der Duin, P., de Ruijter, P. and van Wijck, P. (2010) "Monitoring the future. Building an early warning system for the Dutch Ministry of Justice", Futures, 42: 454–465.

Butter, M., Brandes, F., Keenan, M., Popper, R., Giesecke, S., Rijkers-Defrasne, S., Braun, A. and Crehan, P. (2009) 'Monitoring foresight activities in Europe and the rest of the world', Final Report of the European Foresight Monitoring Network EFMN. EUR 24043 EN. Brussels: European Commission.

Chen, C. (2006). “CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature”. Journal of the American Society for Information Science and Technology, 57(3), 359–377.

Chen,K. and Guan, J (2011). “A bibliometric investigation of research performance in emerging nanobiopharmaceuticals.” *Journal of Informetrics* 5: 233-247

Cozzens, S., Gatchair, S., Kang, J., Kim, KS., Lee, HJ, Ordóñez, G., and Porter, A. (2010). “Emerging technologies: quantitative identifaction and measurement.” *Technology Analysis & Strategic Management* 22(3): 361-376

Csajbok, E., Berhidi, A., Vasas, L., and Schubert, A. (2007). “Hirsch-inex for countries based on essential science indicators data”, *Scientometrics*, 73(1):91-117

Georghiou, L., and Harper, J.C., (2011). “From priority-setting to articulation of demand: Foresight for research and innovation policy and strategy.” *Futures* 43: 243-251

Glänzel, W. and Thijs, B. (2012). “Using ‘core documents’ for detecting and labelling new emerging topics”. Scientometrics, 91(2): 399-416

Hiltunen, E. (2008) “The future sign and its three dimensions”, Futures, 40: 247–60.

Karypis, G. (2006) CLUTO – a clustering toolkit. Available at: http://glaros.dtc.umn.edu/gkhome/cluto/cluto/overview (accessed 30 May 2012).

Konnola, T., Salo, A., Cagnin, C., Carabias, V. and Vilkkumaa, E. (2012) “Facing the future: Scanning, synthesizing and sense-making in horizon scanning”, Science and Public Policy, 39: 222-231

Kostoff , R. N., Rio , J. A., Cortes, H. D., Smith, C., Smith, A., Wagner, C., Leydesdorff, L., Karypis, G., Malpohl, G., and Tshiteya, R. (2007), “Clustering methodologies for identifying country core competencies, Journal of Information Science”, Journal of Information Science 33 (1) : 21–40.

Liu, Y, Navathe, S.B., Civera, J., Dasigi, V., Ram, A., Ciliax, B.J., and Dingledine, R. (2005). “Text Mining Biomedical Literature for Discovering Gene-to-Gene Relationships: A Comparative Study of Algorithms.” IEEE/ACM Transactions on Computational Biology and Bioinformatics 2(1): 62-76

Porter, A.L, Roessner, J.D, Jin, X-Y., and Newman, N.C. (2002). “Measuring national 'emerging technology' capabilities”, Science and Public Policy, 29(3):189-200

Schultz, W. L. (2006) “The cultural contradictions of managing change: Using horizon scanning in an evidence-based policy context”, Foresight, 8 (4): 3–12.

Shibata, N., Kajikawa, Y., Takeda, Y., & Matsushima, K. (2008). "Detecting emerging research fronts based on topological measures in citation networks of scientific publications." Technovation, 28(11): 758–775.

Sutherland, W.J., and Woodroof, H.J. (2009). "The need for environmental horizon scanning", Trends in Ecology and Evolution Vol.24 (10): 523-527

Upham, S.P., and Small, H. (2010). "Emerging research fronts in science and technology: patterns of new knowledge development". Scientometrics 83:15–38

Van Raan, AFJ (2006). "Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups", Scientometrics, 67(3):491-502

Van Rij, V. (2010) 'Joint horizon scanning: identifying common strategic choices and questions for knowledge', *Science and Public Policy*, 37: 7–18.

Yang, Y., Wu, M., and Cui, L. (2012). "Integration of three visualization methods based on co-word analysis." *Scientometrics* 90(2): 659-673

Zhao, Y and Karypis G. (2005). "Hierarchical clustering algorithms for document datasets." Data Mining and Knowledge Discovery, 10(2):141-168.

Zhao, Y and Karypis G. (2004). "Criterion functions for document clustering: Experiments and analysis." Machine Learning, 55(3):311–331

CHAPTER FIVE: CONCLUSIONS

This concluding chapter is broken into three sections: a summary, a discussion on future directions and policy implications.

Summary

In broadening our understanding of the interplay of international scientific collaborations and national scientific systems, this dissertation makes three methodological contributions rooted in social complexity theory

1. A novel methodology of quantitatively determining the core and periphery of science through utilizing a two-tier power analysis
2. A novel framework for understanding bilateral relations between two countries scientific systems (with an introduction of log-odds ratios statistic to analyzing collaborations over time in the field of scientometrics)
3. A novel methodology for detecting pre-emerging S&T capabilities that a country is developing

Taken together, these three contributions show a range of outcomes at the country level that can be deduced through the study of international scientific collaborations.

Determining the Core and Periphery of Science

The methodology of quantitatively determining the core and periphery of science through utilizing a two-tier power analysis was introduced in chapter two. This chapter discovered that the distribution of a country's article output and degree centrality was a two-tier power law commonly found in "rich-club" phenomena and used this discovery to differentiate countries that are in the core and periphery structure of science.

This methodology was demonstrated over five separate scientific and technical disciplines (as defined in Web of Science): Astronomy & Astrophysics, Energy & Fuels, Nanotechnology & Nanoscience, Nutrition, and Oceanography. Each field exhibited the same two-tiered structure that differentiates the core and periphery. The countries that compose the majority of the core across the five disciplines are similar with variations around the edges of the core. Originally, these comprised of a handful of countries in the core of any field of science pre-1992, but the core has now grown in recent time periods to include 29 countries that maintain core membership in at least one discipline.

There are positive and negative implications to this phenomenon. The positive implication is that membership in the core increases the attractiveness of a country's scientific base for collaboration. The negative implication for those countries left in the periphery: they are becoming increasingly isolated from the scientific elite. This isolation is growing as the gap widens in their relative decline in article output, degree centrality and citation impact.

Understanding bilateral relations between two countries scientific systems

The framework for understanding bilateral relations between two countries scientific systems was introduced in chapter three. This chapter built a multi-level and multi-theoretical framework for examining a country's changing pattern of collaboration over time. It develops five statistics that describe the relation between two countries: collaboration percentage, collaboration intensity, citation impact, affinity ratio and odds ratio. The first four statistics have been used in various papers examining international collaboration in sciences but this was the first usage of the odds ratio to the field to the best of the author's knowledge.

The framework was developed using the case study of China's collaboration in Energy & Fuels. Then the framework was demonstrated over five other case studies: two case studies examined changing priorities in the developed world (USA and France), two examined a country's rise from the periphery to the core (South Korea and Chile) and the last case study examined the African continent for insights into the developing world.

The examples of the United States and France highlight how the analysis provides insights into shifting patterns of collaborations. The case study of the United States shows how the US went from collaborations primarily focused on English-speaking countries and France in the pre-1992 time period, to grow with strong ties across a wide-range of countries in the mid-nineties and a rapid rise of collaboration with China and South Korea in the 2000s. The French case study shows a different shifting pattern that goes from the strengthening of European Union ties in the 1990s to strengthening ties to the South Pacific in the 2000s.

The cases of Chile and South Korea show differently paths in which a developing country can join the ranks of developed nations. The Chilean case study shows a developing country entering the core by exploiting its geographic advantage to forge close ties to Europe and the United States. The South Korean shows a different path from the periphery to the core by building a strong relationship with its neighbors and the United States while developing its own indigenous capabilities.

The example of Africa highlights how this methodology may be used by NGOs whose goal is assistance to the developing world by understanding which countries have access to the continent. It showed a continent cleanly split by French- and English-speaking Africa. France's collaborations are exclusive to its former colonies, while the United States is replacing the United Kingdom and other colonial powers as the prime collaborator in the rest of Africa. This case study also highlighted that there is virtually no intra-African collaboration.

Detecting pre-emerging S&T capabilities that a country is developing

The methodology for detecting pre-emerging S&T capabilities that a country is developing was introduced in chapter four. This chapter combines the horizon scanning methodology developed by future studies with emergence detection algorithms develop by information scientists to develop the Quantitative Exploratory Horizon Scan for identifying emerging country capabilities that may provide it a distinct advantage. This methodology involves clusters a country's scientific output and seeks to identify small clusters that are experiencing rapid growth with a low rate of international collaboration (defined as its purity). By combining these three statistics, clusters can be ranked by those most likely to become scientific areas in which the country will have a distinct advantage in the near term.

This chapter applied this methodology to a case study of the United States' scientific output from 2005 to 2010 and identified emerging areas in biology, chemistry and physics in which the United States has significant advantages in. Several of these areas are related to areas in which recent Nobel Prizes were awarded in. An analysis of two emerging areas, Optical Frequency Combs and the gene TH-17, was conducted to verify and validate that this methodology.

Future Research Directions

Throughout the development of the various methodologies in this dissertation, case studies were used to highlight the application of each technique. In chapters two and three, an attempt was made to select five diverse case studies to understand how these methods fared under different scientific and technological disciplines. While diverse, these studies relied on the subject categorization of the Web of Science. While this made data collection and repeatability easier, there may be biases built into this collection that need to be explored to determine how robust these methodologies are. This exploration may be accomplished through either exploiting different data sources, such as Scopus or IEEE proceedings, or by building user-defined queries to define technological areas. Additionally, there is likely some sensitivity to the size and/or maturity of a scientific or technological area to these methodologies that needs to be explored.

The Quantitative Exploratory Horizon Scan developed in chapter four is by far the most experimental methodology developed within this dissertation and would benefit most by additional case studies as well as addition validation and verification. There are also several ways this methodology can be extended to performing horizon scans for pre-emergence of an institution's capabilities and subtopics in broad subject areas such as computer science, material science or neuroscience.

It is worth reiterating that the original scope of this dissertation was in the application of social complexity methodologies to quantitatively study national scientific systems through their interaction with other nations. Thus the scope of this research was to uncover trends without explaining the reason for these trends. Indeed, most of the empirical work from chapters two through four raise more questions than they answer. Did South Korea actively pursue a policy to tie them to the United States? Did the UK intentionally cede English-speaking Africa to the United States? What mechanisms did the European Union use to increase collaboration within the EU in the 1990s? Are the technologies found in the Quantitative Exploratory Horizon Scan going to give the United States a lasting technological lead? What forces are causing China to turn away from Europe just when Europe is trying to increase its collaboration with China?

It is beyond the scope of this dissertation to answer such questions, even though answering such questions would validate the methodologies proposed within this dissertation. Answering such questions will require a mixed methodological approach utilizing case studies, interviews and/or partnering with subject matter experts and statistics from a variety of other data sources. This research is left for future endeavors.

Policy Implication of this Research

This research was motivated by the desire to provide a new set of tools that can be used for understanding the effect of emergent networks of collaboration in today's global scientific community and how external factors can shape them. This research has built upon a growing body of literature seeking to understand the dynamics of international scientific collaboration and its interaction on national systems. It is my hope that these findings and tools will be informative to policy-makers for developing strategies to best position their respective scientific enterprises in today's globalized world. There are three policy areas worth noting that this research can be used to assist policy makers: policy planning, policy evaluation, and science and technological intelligence gathering.

In the policy planning context, increasing international cooperation is a goal of policy makers. Understanding the current trends in collaboration between two countries allows policy makers to specifically target policies to either increase or decrease relations between two countries. Identifying trends as they are happening, such as the decrease of US-Russia collaboration in Energy or the withdrawal of the UK from Africa, would enable planners to tune their policies to support or counteract the trends.

Utilizing this research for planning for development, understanding just how far into a periphery of science helps determine the level of effort needed to bring a country up to the core. The trend identified in chapter 2 of the periphery becoming increasingly isolated from the core is troubling. Understanding the linkages between the developing and developed world can show which countries have the highest probability of success working with the developing country to help them bridge this increasing divide. The case studies of South Korea and Chile provide examples where this has been successful. Chile's example shows how a country can exploit a geographical advantage to catapult their system in a selective discipline; while South Korea provides a more generalized path of how strong collaborative ties with targeted nations can help. Though the case study on Africa provides a counter-example of how strong ties to only a single country is not enough to lift it out of the periphery. Understanding these case studies can provide invaluable assistance for policy planning.

These tools are also important for evaluating the outcomes of policies that are aimed at strengthening scientific relations between countries. As shown in chapter three, after the fall of the Soviet Union, the United States and Russia had a sharp increase in collaboration that was not sustained in the 2000s. One can look at efforts launched to increase ties between the two countries during that time period and see which of these, if any, were responsible for this outcome. Additionally, looking at long term policies between the United States and South Korea may also be studied to determine if the results identified in chapters two and three were the cause of any particular policies to lift South Korea from the periphery of science to the core.

Lastly, the importance of science and technological intelligence gathering should not go unmentioned. Advanced countries cannot maintain advantage across all areas of science. Chapter four provides methods for understanding the emerging capabilities of countries that can decrease the chance of technical surprise and inform policy makers of areas that they should be investing in within their own borders. It may also inform policy makers of technological areas within their borders that can have a high potential for growth and thus could be targeted for action to strengthen these fields.

The common theme here is that these tools contribute to the understanding and evaluation of the dynamics of international scientific collaboration thus allowing policy makers and policy analysts to be better informed and increase their ability to react in a positive manner to the quickly changing global scientific landscape.

REFERENCES

Ahlqvist T, Valovirta V, and Loikkanen T (2012). "Innovation policy roadmapping as a systemic instrument for forward-looking policy design", *Science and Public Policy* 39: 178–190

Amanatidou, E., Butter, M., Carabias, V., Konnola, T., Leis, M., Saritas, O., Schaper-Rinkel, P. and van Rij, V. (2012) 'On concepts and methods in horizon scanning: Lessons from initiating policy dialogues on emerging issues', *Science and Public Policy*, 39: 208–21.

Amaral, L.A.N. & J.M. Ottino (2004). "Complex networks: Augmenting the framework for the study of complex systems". *Eur. Phys. J. B* 38, 147–162

Ansoff, H. I. (1975) "Managing strategic surprise by response to weak signals", *California Management Review*, 18: 21–33.

Barabasi, A.L., Jeong, H.,N´eda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica A* 311, 590–614.

Beaver, D. deB., Rosen R. (1979), "Studies in scientific collaboration. Pt 2. Scientific co-authorship, research productivity and visibility in the French scientific elite, 1799-1830", *Scientometrics* 1 133-149

Ben-David J. (1971). *The Scientist's Role in Society*. Englewood Cliffs, NJ: Prentice-Hall.

Bettencourt, L.M.A. , Kaiser, D.I. , Kaur, J., Castillo-Ch´vez, C. and Wojick, D.E. (2008). "Population modeling of the emergence and development of scientific fields", *Scientometrics*, 75: 495–518.

Borgatti, S.P., and Everett, M.G. (1999). "Models of Core/Periphery Structures." *Social Networks* 21, 375–395

Botterhuis, L., van der Duin, P., de Ruijter, P. and van Wijck, P. (2010) "Monitoring the future. Building an early warning system for the Dutch Ministry of Justice", *Futures*, 42: 454–465.

Boyack, K. W., & Klavans, R. (2010). “Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately?” *Journal of the American Society for Information Science and Technology*, 61(12), 2389–2404.

Butter, M., Brandes, F., Keenan, M., Popper, R., Giesecke, S., Rijkers-Defrasne, S., Braun, A. and Crehan, P. (2009) ‘Monitoring foresight activities in Europe and the rest of the world’, *Final Report of the European Foresight Monitoring Network EFMN*. EUR 24043 EN. Brussels: European Commission.

Callon, M., J. Law, A. Rip, (1986), *Mapping the Dynamics of science and technology: Sociology of Science in the read world* (Macmillam, Houndmills, U.K.)

Chen, C. (2006). “CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature”. *Journal of the American Society for Information Science and Technology*, 57(3), 359–377.

Chen,K. and Guan, J (2011). “A bibliometric investigation of research performance in emerging nanobiopharmaceuticals.” *Journal of Informetrics* 5: 233-247

Cioffi-Revilla, C. (eds.) (forthcoming). *Power Laws in the Social Sciences: Discovering Complexity and Non-Equilibrium Dynamics in the Social Universe*. Publication forthcoming.

Cozzens, S., Gatchair, S., Kang, J., Kim, KS., Lee, HJ, Ordóñez, G., and Porter, A. (2010). “Emerging technologies: quantitative identifaction and measurement.” *Technology Analysis & Strategic Management* 22(3): 361-376

Csajbok, E., Berhidi, A., Vasas, L., and Schubert, A. (2007). “Hirsch-inex for countries based on essential science indicators data”, *Scientometrics*, 73(1):91-117

De Solla, D.J. (1965) Networks of Scientific Papers. *Science* 149 (3683), 510-515

Drori, G.S., Meyer, J.W., Ramirez, F.O., and Schofer, E., (2003), *Science in the Modern World Polity: Institutionalization and Globalization* (Stanford University Press, Stanford CA)

Duerbeck, HW. “National and international astronomical activities in Chile 1849-2002” in *Interplay of Periodic, Cyclic and Stochastic Variability in Selected Areas of the H-R Diagram*. Edited by C. Sterken, ASP Conf. Ser. 292. San Francisco: Astronomical Society of the Pacific, 2003, p. 3-18.

Gauffiau, M., P.O. Larsen, I. Maye, A. Roulin-Perriard, & M. Von Ins (2007). "Publication, cooperation and productivity measures in scientific research." *Scientometrics*, Vol. 73(2), 175–214

Georghiou, L. (1998). Global cooperation in research. *Research Policy* 27, 611–626.

Georghiou, L., and Harper, J.C., (2011). "From priority-setting to articulation of demand: Foresight for research and innovation policy and strategy." *Futures* 43: 243-251

Glänzel W. (2000), "Science in Scandinavia: A bibliometric approach", *Scientometrics* 48 121–150. (Correction: Scientometrics, 49 (2000) 357.

Glänzel, W. (2001). "National characteristics in international scientific co-authorship relations." *Scientometrics*, Vol. 51(1), 69–115

Glänzel W., Schubert A., (2001) "Double effort = double impact? A critical view at international co-authorship in chemistry", Scientometrics, 50 199–214.

Glänzel, W., and Schubert A. (2004). "Analyzing scientific networks through co-authorship." In: Moed, H.F., Glänzel, W., and Schmoch, U. (2004) (eds.) *Handbook of quantitative science and technology research. The use of publication and patent statistis in studies of S&T systems* (pp. 257-276). Dordrecht (the Netherlands): Kluwer Academic Press

Glänzel, W., and Schubert A. (2005). "Domesticity and internationality in co-authorship, references and citations." *Scientometrics* 65(3), 323-342

Glänzel, W., Debackere, K., and Mayer, M. (2008). "'Triad' or 'tetrad'? On Global changes in a dynamic world." *Scientometrics*, Vol. 74(1), 71-88

Glänzel, W. and Thijs, B. (2012). "Using 'core documents' for detecting and labelling new emerging topics". *Scientometrics*, 91(2): 399-416

Hill, C.T. (2007). "The Post-Scientific Society." *Issues in Science & Technology*, 24(1)

Hiltunen, E. (2008) "The future sign and its three dimensions", *Futures*, 40: 247–60.

Holland, J. (1998). *Emergence: From Chaos To Order*. New York: Helix Books

Hwang, K. (2008) "International Collaboration in Multilayered Center-Periphery in the Globalization of Science and Technology", *Science Technology Human Values* 33(1); 101-133

Jeong, H., Neda,A., & Barabasi, A.L. (2001). “Measuring preferential attachment for evolving networks.” arXiv:cond-mat/0104131 v1, 7 April 2001.

Karypis, G. (2006) CLUTO – a clustering toolkit. Available at: http://glaros.dtc.umn.edu/gkhome/cluto/cluto/overview (accessed 30 May 2012).

Katz, J.S., & Hicks, D. (1997). “How much is a collaboration worth? A calibrated bibliometric model.” *Scientometrics* 40 (3), 541–554.

Katz, J.S., & Martin, B. R. (1997). „What is research collaboration?” *Research Policy,* 26(1), 1–18.

King, DA. (2004). “The Scientific Impact of Nations.” *Nature* V432, 311-316

Konnola, T., Salo, A., Cagnin, C., Carabias, V. and Vilkkumaa, E. (2012) “Facing the future: Scanning, synthesizing and sense-making in horizon scanning”, *Science and Public Policy*, 39: 222-231

Kostoff , R. N., Rio , J. A., Cortes, H. D., Smith, C., Smith, A., Wagner, C., Leydesdorff, L., Karypis, G., Malpohl, G., and Tshiteya, R. (2007), “Clustering methodologies for identifying country core competencies, Journal of Information Science”, *Journal of Information Science* 33 (1) : 21–40.

Laudel, G. (2001). “Collaboration, creativity and rewards: why and how scientists collaborate?” *International Journal of Technology Management*, 22 (7/8), 762-781.

Lall, S. (2001). “Competitiveness indices and developing countries: an economic evaluation of the Global Competitiveness Report”, *World Development*, 29(9), 1501-1525.

Latour, B. and S. Woolgar, (1986). *Laboratory Life: The construction of social facts, 2nd Edition* (Princeton University Press, Princeton)

Leydesdorff, L., & C.S. Wagner (2008). “International Collaboration in Science and the Formation of a Core Group**.”** *Journal of Informetrics* 2(4), 317-325

Leydesdorff L, Wagner CS, “Is the United States losing ground in science? A global perspective on the world science system.” *Scientometrics*, Vol 78, No. 1 (2009) 23-36

Liu, Y, Navathe, S.B., Civera, J., Dasigi, V., Ram, A., Ciliax, B.J., and Dingledine, R. (2005). “Text Mining Biomedical Literature for Discovering Gene-to-Gene Relationships: A Comparative Study of Algorithms.” *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 2(1): 62-76

Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences* **16** (12), 317–324.

Luukkonen T., Persson O., Sivertsen G. (1992), "Understanding Patterns of International Scientific Collaboration", Science*, Technology, & Human Values*, 17(1) 101-126

Merton, R.K. (1968). "The Matthew Effect in Science The reward and communication systems of science are considered**,"** Science 159 (3810) 56

Merton, R.K., (1973). *The Sociology of Science: Theoretical and Empirical Investigations.* Chicago: University of Chicago Press

Moed H. F. (2005). "Does International Collaboration Pay?" in *Citation Analysis in Research Evaluation*. Springer: The Netherlands

Moed H. F., De Bruin R. E. and Van Leeuwen T. N. (1995) "New bibliometric tools for the assessment of national research performance: database description, overview of indicators and first applications." *Scientometrics*, 33: 381–422

Monge, P. R., & Contractor, N. (2003). *Theories of Communication Networks*. New York: Oxford University Press

Newman, M.E.J. (2001). "The structure of scientific collaboration networks." *Proceedings of the National Academy of Sciences* 98, 404–409.

Newman, M.E.J., (2004). "Coauthorship networks and patterns of scientific collaboration." *Proceedings of the National Academy of Sciences* 101 (Suppl. 1), 5200–5205.

Ohmae, K. (1985). *Triad Power: The Coming Shape of Global Competition*. New York: Free Press

Oldham, G. (2005). *Policy Brief: International scientific collaboration: a quick guide, Science and Development Network*, April 2005

Porter, A.L, Roessner, J.D, Jin, X-Y., and Newman, N.C. (2002). "Measuring national 'emerging technology' capabilities", Science and Public Policy, 29(3):189-200

Schofer E., Ramirez F.O., Meyer J.W., (2008) "The effects of science on national economic development, 1970 to 1990", *American Sociological Review* 65 (6): 866-887

Schott, T. (1993). "World science: globalization of institutions and participation." *Science, Technology and Human Values*, 18, 196–208

Schott, T. (1998). “Ties between center and periphery in the scientific world system.” *Journal of World-Systems Research* 4, 112-144

Schott T., “Global Webs of Knowledge: Education, Science, and Technology”, *American Behavioral Scientist*, Vol 44, No. 10 (2001) 1740-1751

Schubert A. and Braun T, (1986).”Relative indicators and relational charts for comparative-assessment of publication output and citation impact”. Scientometrics, 9: 281–291.

Schultz, W. L. (2006) “The cultural contradictions of managing change: Using horizon scanning in an evidence-based policy context”, *Foresight*, 8 (4): 3–12.

Serrano, M.A. (2008). Rich-club vs rich-multipolarization phenomena in weighted networks. *Physical Review E* 78, 026101

Shibata, N., Kajikawa, Y., Takeda, Y., & Matsushima, K. (2008). “Detecting emerging research fronts based on topological measures in citation networks of scientific publications.” *Technovation*, 28(11): 758–775.

Smith, M. (1958), “The trend toward multiple authorship in psychology”, *American Psychologist* 13, 596-599.

Stefaniak, B. (2001). “International co-operation in science and in social sciences as reflected in multinational papers indexed in SCI and SSCI”, *Scientometrics* Vol. 52(2) 193–210

Sutherland, W.J., and Woodroof, H.J. (2009). “The need for environmental horizon scanning”, *Trends in Ecology and Evolution* Vol.24 (10): 523-527

Thomson, “The Thomson Scientific Journal Selection Process”, accessed on website 6/12/2010: http://www.thomsonreuters.com/business_units/scientific/free/essays/journalselection/

Traweek, S. (1988). *Beamtimes and lifetimes: The world of high energy physicists*. Cambridge, MA:Harvard University Press.

United Nations Standing Committee on Nutrition.(2009) “Why Nutrition is important.” *United Nations Standing Committee on Nutrition*. Web. 18 Apr. 2012

Upham, S.P., and Small, H. (2010). “Emerging research fronts in science and technology: patterns of new knowledge development”. *Scientometrics* 83:15–38

Van Raan, AFJ (2006). "Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups", *Scientometrics*, 67(3):491-502

Van Rij, V. (2010) 'Joint horizon scanning: identifying common strategic choices and questions for knowledge', *Science and Public Policy*, 37: 7–18.

Wagner, C.S. (1997) *International Cooperation in Research and Development: An Inventory of U.S. Government Spending and a Framework for Measuring Benefits*, Santa Monica, CA: Rand.

Wagner C.S. (2005). "Six case studies of international collaboration in science." *Scientometrics* 62 (1), 3-26

Wagner, CS (2008). *The New Invisible College: Science for Development*. Washington DC: Brookings Institution Press

Wagner C.S., Leydesdorff L, (2005). "Network structure, self-organization, and the growth of international collaboration in science." *Research Policy* 34 (10), p.1608-1618

Wagner-Döbler, R., (2001). "Continuity and discontinuity of collaboration behaviour since 1800 – from a bibliometric point of view." *Scientometrics* 52, 503–517.

Yang, Y., Wu, M., and Cui, L. (2012). "Integration of three visualization methods based on co-word analysis." *Scientometrics* 90(2): 659-673

Zhao, Y and Karypis G. (2005). "Hierarchical clustering algorithms for document datasets." *Data Mining and Knowledge Discovery*, 10(2):141-168.

Zhao, Y and Karypis G. (2004). "Criterion functions for document clustering: Experiments and analysis." *Machine Learning*, 55(3):311–331

Zelnio, R. (2011), "Structure of International Scientific Collaboration." In: E. Noyons, P. Ngulube, J. Leta (Eds), *Proceedings of ISSI 2011 - The 13th International Conference on Scientometrics and Informetrics*, Durban, South Africa, 4-7 July 2011, 898-913.

Zelnio, R.(2012), "Identifying the global core-periphery structure of science." *Scientometrics*, Vol.91, No. 2 (2012) 601-615

Zhou, S. and R. J. Mondragón (2004). "The Rich-Club Phenomenon in the Internet Topology." *IEEE Communication Letters* 8(3), 180-1

CURRICULUM VITAE

Ryan J. Zelnio graduated from Bettendorf High School, Bettendorf, Iowa, in 1992. He received his Bachelor of Science in Computer Science and Bachelor of Science in Mathematics from Marycrest International University in 1998. He was employed as a software engineer and a manager for a variety of companies from 1995 to 2006 and received his Master of Science in Space Studies from University of North Dakota in 2004. He currently is a Lead Scientist for the United States Navy at the Naval Surface Warfare Center – Dahlgren Division.