Research Collaboration, Academic Stars and the Evolution of Science Systems.

A dissertation submitted in partial fulfillment of the requirement of the degree of Doctor of Philosophy in Engineering and Public Policy

Leonardo Reyes-Gonzalez

May 2012

Carnegie Mellon University
Carnegie Institute of Technology
Department of Engineering and Public Policy
Program in Strategy Entrepreneurship and Technological Change
Pittsburgh, Pennsylvania 15213 USA

Permanent contact E-mail: ideado.MX [at] gmail [dot] com

To my family:

(1) the one life gave me

(Juanita, Jesus, Ivan and Mariana),

(2) the one I'm creating with Josune,

and (3) Lola and Lolek for cheering and chirping my life.

Contents

List of Figures

.

List of Tables

Acknowledgements

Over the last years I have been blessed with the support and friendship of a myriad of people. In addition, several more have influenced my research and directly or indirectly aided in completion of this dissertation. First, I would like to acknowledge the contribution of Francisco Veloso, my advisor and Chair of my thesis committee. Without Francisco's support, encouragement and patience this work would not exist. Thank you for showing me what it means to collaborate with an eminent person. I also want to thank Elizabeth Casman (member of my thesis committee), Beatrice Dias and Claudia Gonzalez-Brambila (member of my thesis committee) for showing me what it really means to do research in a nurturing environment. In addition, I want to thank Sara Kiesler (member of my thesis committee) for reviewing my work and giving me valuable feedback. Furthermore, I want to thank Miguel Angel Perez-Angon for encouraging me to study the area of physics in Mexico and helping me test my initial thoughts in this field; and Margarita Calleja-Quevedo for our long conversations on science, technology and innovation policy in Mexico, as well Mexican politics. Finally, I want to thank Granger Morgan (member of my thesis committee) for helpful advice and for facilitating the interdisciplinary study of engineering and public policy problems.

This work was supported by three grants from CMU's Berkman Faculty Development Fund; the Mexican Council for Science and Technology (CONACYT) fellowship 168868, and National Science Foundation grant NSF-SBE-0738182. Francisco deserves special recognition (and gratitude) for his efforts to secure the financial support for this research and for my Ph.D. studies.

I want to acknowledge the supporting effort of Bill McDowell who helped improve the algorithm in Chapter 2 and developed the software so we could find the relevant units of analysis automatically. I owe you special thanks for your hard work. In addition, I want to thank Janette Lever for her invaluable support in proof reading this work and helping me improve my English.

Outside "formal" academia there were great people at Carnegie Mellon University that showed me why this institution is an outstanding place beyond its research. First, I want to thank Indira Nair, Suzie Laurich-McIntyre, Nancy Klancher, Lisa Krieg, Linda Gentile, Jennifer McNabb, and Neslihan Ozdoganlar for inviting me to work with them and allowing me to serve the graduate student community. In addition, I want to thank Patti Steranchak, Vicki Finney, Barb Bugosh, Dee Matthews and all the staff at EPP for their friendliness and helping me deal with the red tape.

A special thanks goes out to Matt Hamilton for being my colleague at school and my buddy in Pittsburgh. Thank you for all your support and friendship! I also want to thank Julia Steitz, Paulina Jaramillo, Alexander Mateus, Ines Azevedo, Jennifer Logue and Elizabeth Gilmore for all their encouragement and for making Pittsburgh a great place to live; and acknowledge the contribution in this respect to Rahul Walawalkar, Janice Tsai, Paul Enders, Tamar Krishnamurti, Guillermo Jose-Aguilar, Monica Gonzalez-Veron, Carla Leininger, Mauren Antkowski, and Pedro Leon. In addition, I want to thank Lee Vinsel, Faheem Hussain and Chirantan Chatterjee for their endless discussion of science, technology, policy and economics; and letting me see that life goes beyond this issues. Last but not least I want to thank Miriam Esquitin for our long enduring friendship, and her long distance support!

Finally, I want to thank my family for their endless love and support. I want to thank my Mother, Juanita, for always being there for me and showing me throughout my entire life the true meaning of expression “never, never, never, give up…” no matter what life gives you or takes away form you. I want to thank my brother for his support and balancing my mind trough endless talks about photography, design and creativity; and my sister for all her love. I want to thank Giuseppe, Rita and Nino; as well as Hilda, Vale, Tarini and Chuy for all their encouragement and letting me be part of your family. Most importantly, I owe a lifetime of thanks to my wife, Josune, for believing in me and for your patience, your support, your understanding, your friendship, your love, for everything. Thank you and I love you.

Abstract

The important growth in research collaboration is generating increasing attention by research administrators and policy makers. There is much interest in improving our understanding of the nature, dynamics and impact of this cooperation in science. This thesis contributes to this area in three dimensions. First, it proposes a novel method by which one can characterize and assess research collaboration, which takes into consideration the self-organizing process of scientific collaboration. Second, building partially on the new method, it studies how research collaboration, in particular research groups and scientific stars, influence the nurturing of new researchers that enter a scientific system. Finally, it explores in detail what the new researchers look for, and find, in their early collaborations. The field of physics and related areas (including applied physics, material sciences and optics) in Mexico is used to look at these issues.

The proposed evaluation method uses self-organizing characteristics of science to identify and compare relevant units of analysis. To characterize groups, the thesis exploits the patterns of collaboration and develops a method that identifies and benchmarks research groups. Collaboration patterns of researchers are used to identify the frontiers of the focal research units and the backward citation patterns are employed to establish relevant benchmark units for each focal unit. The results suggest that the definition of the unit of analysis affects our understanding of the position a research institutions has within the Science Technology and Innovation (ST&I) System and provides evidence that the performance of Mexican institutions in Physics is highly heterogeneous within institutions. This is important because research administrators and policy makers need to take into account this heterogeneity when assessing the ST&I system.

The second contribution of this thesis is an investigation of how different forms of scientific collaboration early on in the career of a researcher relate to his or her future publication and citation rates, and their likelihood of becoming a leading scientist. In particular it quantifies the effect of collaborative research environments, such as prominent scientists or research groups (identified using the method developed in the thesis), on new scholars. This study shows that eminent scientists have an important role in the development of a scientific system (especially within the context of an emerging economy) in terms of publications and citations. In particular it finds that these stars have a positive and significant effect on the productivity and impact of young researchers, as well as on their likelihood of also becoming leading scientists. In addition, early collaboration with a highly productive research group and the leader of this group also contributes to superior productivity performance by scientists.

Third, this thesis explores how budding scientists, some of which became highly accomplished researchers, used their collaborations with other top scientists and research groups early in their career. This works finds that researchers who later became star scientists focus on acquiring new ideas and knowledge through early interactions with other scientists, particularly foreign collaborators and existing stars, whereas those less prominent focus on accessing resources and only learning “basic” research kills, like publishing.

Finally, this thesis provides important insights for policy makers by showing the significance research collaboration has in the development of ST&I of an emerging economy. In addition, this work highlights the importance of endogenously defining the unit of analysis and taking into account the heterogeneity within the system when making assessments of the ST&I system. Furthermore, this dissertation shows the relevance scientific stars surrounded by nurturing environments have in the progress of science, as well as the importance cooperation with these scientists and foreign collaboration has in exposing young faculty to new ideas.

Chapter 1. Introduction

Today's emphasis on economic activity based on knowledge and innovation is leading industrialized as well as developing nations to place special attention on their science, technology and innovation (ST&I) systems (OECD, 2004b, 2011). In particular, there is strong emphasis in promoting policies that foster its progress and enhance its economic and social benefits (OECD, 1999, 2001, 2004a, 2004b, 2011). At the core of this system lies the individual researcher – in particular highly productive ones (Lotka, 1926; Pirce, 1963; Hagstrom, 1968; Cole, 1970; Cole and Cole, 1972; Allison and Stewart, 1974; Reskin, 1977, 1978; Fox, 1983; Zucker et al., 1998; Azoulay et al., 2007, Goodall, 2009; Oettl, 2009). They are responsible for expanding our knowledge base (Kuhn, 1962), transferring this knowledge into useful applications (Gibbons et al., 1994; Jain et al., 2009; Ding and Choi, 2011), and training and coaching the next generation of scientists and technologists (Bozeman and Corley, 2004; Ham and Weinberg, 2007; Waldinger, 2010).

In addition to the individual, the group of research collaborators, or teams, is another dimension considered to be of paramount importance in the development of a ST&I system (Adams, et al., 2005; Wagner and Leydesdorff 2005; Wuchty et al., 2007). It is within a team that researchers can acquire new skills (Wagner, 2008, p2); and gain access to complementary expertise (Katz and Martin, 1997; Melin, 2000; Beaver, 2001), as well as valuable equipment and resources (Melin, 2000; Beaver, 2001). Research teams are also the primary environment where scientists are exposed to new ideas (Katz and Martin, 1997; Melin, 2000) and where they can increase their visibility and prestige (Crane, 1972; Beaver and Rosen, 1978, 1979a,b; Katz and Martin, 1997; Beaver, 2001). Furthermore, this is the setting through which cross-fertilization across fields is established (Beaver and Rosen, 1978, 1979a, 1979b; Katz and Martin, 1997; Melin, 2000).

These observations have motivated an important stream of research focused on understanding the role scientific cooperation plays in the development of ST&I systems. In particular, a great effort has been dedicated to assessing the effect research collaboration and the different environments that emerge from these interactions, like conducting research with a star-scientist[1], has on the performance of researchers and the progress of their academic careers. In addition, previous research has also looked at how these connections in science can be used to identify the different invisible colleges in ST&I systems. It is within this area of study that researchers have used these interactions, and their occurrence and the patterns that emerge from them to developed algorithms that identify communities within co-authorship networks. In the remaining part of this chapter the literature that supports this thesis is reviewed. Additionally, the studies conducted in this work are summarized and their main findings highlighted. This section ends laying the structure of this dissertation.

1.1. Research Collaboration

The interactions that happen between scholars and the groups that emerge from these exchanges plays an important role in the development of the ST&I system. A variety of studies have looked at how research collaboration influences the progress of science. According to Bukvova (2010) these studies have primarily focused on five main domains: (a) defining what is research collaboration; (b) problems with measuring this phenomenon; (c) understanding why

---

[1] In this work the terms *star, highly regarded, eminent, prominent or key researcher/scientist* are used interchangeably to refer to individuals that have (or are perceived of having) on average a higher performance than their peers.

researchers collaborate; (d) developing explanatory approaches to research collaborations; and, in recent years, (e) understanding the role Information and Communication Technologies (ICT) plays in promoting this type of activities in science. In particular, there has been special interest in defining what is research collaboration and developing appropriate methods to measures these interactions (Melin & Persson, 1996; Katz and Martin, 1997; Laudel, 2002); as well as, in uncovering the factors that promote and hinder cooperation in science (Bukvova's, 2010). Similarly, previous studies have tried to assess the costs, benefits and opportunities these schemes produce to individual researchers throughout their professional career (Katz and Martin, 1997; deB. Beaver, 2001). In the following sections we review these studies.

1.2.1. What is Research Collaboration and how it can be Measured

Research collaboration is not easy to define and measure because it is "largely a matter of social convention among scientists" that varies "across institutions, fields, sectors and countries" and changes over time (Katz and Martin, 1997). Collaboration takes place between individuals who primarily are researchers and belong to one or more institutions from one or more regions of the world. These interactions happen within and across fields of knowledge and, in some same cases, this cooperation happens at different organizational levels, such as departments or institutions (Bukvova, 2010). Some authors have tried to define research collaboration explicitly. For example, Laudel (1999, p.32; 2002) defines research collaboration as a "system of research activities by several actors related in a functional way and coordinated to attain a research goal corresponding with these actors' research goals or interests." Prior work has used a variety of dimensions to describe it, including the professional background and institutional affiliation of the participants, their disciplinary focus, their geographical location and the organizational level where these interactions happen (Amabile et al., 2001; Sonnenwald, 2007)[2]. One way that has been extensively used to characterize cooperation in science is through bibliometric indicators (Laudal, 2002), in particular co-authorship. This provides many opportunities, although also some important limitations. For example, collaboration through co-authorship (i.e. multi-author or multi-address papers) is a partial indicator of this activity because sometimes collaboration does not lead to a co-authored paper. Conversely, peripheral interaction between scientists can yield co-authored publications (Katz and Martin; 1997). Studies have acknowledged (to a certain degree) the limitations stated by Katz and his colleague, but typically defend the use of co-authorship as a measure of collaboration, arguing that this type of data is the most objective, tangible and well-documented form of scientific collaboration (Newman, 2001; Glänzel and Schubert, 2004). Five key advantages are particularly relevant: verifiability, stability over time, reduced cost, unobtrusiveness, and ease of measurement (Katz and Martin, 1997; He et al., 2009).

1.2.2. Factors Affecting Research Collaboration

The scientific community has dedicated a great effort to the understanding of the factors that promote, enhance and hinder collaboration in science. These range from individual characteristics, such as particular personalities being suited for collaborative work (Stokols et al., 2008), to group attributes, like size (Rigby, 2009), and ability to coordinate (Cummings and Kiesler, 2007), communicate (Stokols et al., 2008) and deal with differences (Jeffrey, 2003; Bammer 2008). Also important are institutional features, like academic culture (Sorensen, 2003)

---

[2] Laudal (2002) indentifies two forms of specialization: *Vertical specialization* observed in theoretical and conceptual activities (like the ones found in apprenticeship, teacher/student or group-leader/doctoral student relationships) and *horizontal specialization* occurring at both levels of vertical specialization (e.g. group-leaders or doctoral students).

or granting scientific credit (Kennedy, 2003; Birnholtz, 2008) and policy aspects, such as funding (Defazio et al., 2009) or national security (Dias et al., 2010).

1.2.3. Benefits and Costs of Research Collaboration

Many authors have looked at the benefits and costs of scientific collaborations to individual researchers throughout their career. Cooperation in science can generate many advantages, including access to expertise, funding and resources (like instrumentation and data sets), exchange of ideas (especially across disciplines), learning new skills, pooling expertise for complex problems, prestige, and in some cases fun and pleasure (Katz & Martin, 1997; deB Beaver, 2001; Bukvova, 2010). However, collaboration is not risk free activity. In some instances, this endeavor can produce financial costs (from traveling or relocation), increase bureaucratic and managerial costs, in particular coordination costs (Cummings and Kiesler, 2007). It can also be difficult to reconcile organizational and cultural differences (Katz and Martin, 1997) and in some instance it can be difficult to assign credit to the participants (Wray, 2006).

The research in this dissertation leverages the use of collaboration to, first, develop and test a method for the characterization and assessment of scientific teams. Then, it uses these interactions to quantify the impact early collaboration with eminent scientists and other forms of nurturing environments have on the performance and professional development of young faculty. Finally, this thesis also uses this early cooperation with different research environments to qualitatively assess the benefits of research collaboration.

1.3. Eminent Scientists

In recent times, there has been a growing interest among research administrators, policy makers and scholars on the role eminent scientists have on the development of ST&I system. This interest stems from the fact that this group of people is responsible for many groundbreaking discoveries (Kuhn, 1962), produce a disproportional amount of the research (Lotka, 1926; Pirce, 1963; Zucker et al., 1998) and receive a large share of citations (Hagstrom, 1968; Cole, 1970; Cole and Cole, 1972; Allison and Stewart, 1974). For example, Price (1963) found that six percent of physicists were responsible for publishing 50 percent of all the publications, while Cole (1979) and Reskin (1977,1978) have shown that this percentage of contributing scientists is 2.5 times higher in other fields. Additionally, Zuker and Darby (1998) discovered that 0.8% of the scientists contributing to the GenBank[3] were 22 times more productive than the average scientist, publishing 17.3% more papers. Furthermore, Allison and Stewart (1974) also found that the distribution of citations is more unequal than the one for articles and that this inequality increases with tenure, for both measures.

Star scientists are also important to research systems because they are responsible for training and coaching the next generation of highly qualified personnel. For instance, Zuckerman (1967) showed that 62% of the Noble laureates (in his sample) worked as young researchers under the supervision of previous prizewinners and Ham and Weinberg (2007) proved that being surrounded by other prizewinners had a significant positive effect on starting their own work that would yield this type of recognition. Furthermore, there is evidence that prominent scientists are

[3] GenBank is a sequence database, produced and maintained by the National Center for Biotechnology Information (NCBI), which is part of part of the National Institutes of Health (NIH) in the United States.

more inclined to collaborate with other distinguished and highly productive researchers than with their less renowned counterpart (Zuckerman, 1967).

Empirical studies have also shown that lead scientists have a positive effect on the productivity of their collaborators. For example, Azoulay et al., (2010) as well as Oettl (2009) have measured the impact that "superstars" have on their peers by calculating the drop in productivity when a leading scientist ceases to exist. In the first study, Azoulay and his colleagues show that coauthors of an 'extinguished' star "suffer a lasting 8 to 18% decline in their quality-adjusted publication output;" whereas Oettl documents a higher loss in productivity, between 19 to 35%. In addition, Waldinger (2010), using a similar exogenous variation (in this case the expulsion of mathematics professors in Nazi Germany), shows that faculty excellence is a very important determinant of short- and long-run doctoral student performance; according to him

> one-standard-deviation increase in faculty quality increases the probability of publishing the dissertation in a top journal by 13 percentage points, the probability of becoming a full professor by 10 percentage points, the probability of having positive lifetime citations by 16 percentage points, and the number of lifetime citations by 6.3.

Star-scientists also matter as role models for the next generation of researchers, where the later tend to emulate the steps of their advisors and collaborators. As previously stated, young scientists working with Nobel Laureates have a higher propensity of replicating the success of their senior collaborators (Zuckerman, 1967). In addition, the mentorship literature has noted that "the majority of participating mentors had been involved in a previous mentoring relationship as a protégé" (Allen el al., 1997) and showed that protégés that were trained by high fecundity[4] mentors also score high on this indicator (Malmgren et al., (2010). This suggests that advisees follow the steps of their advisors in terms of creating mentoring environments once they become research leaders.

Finally, star-scientists are also important in other settings like research administrators or entrepreneurs. For example, Goodall (2009) shows that research quality of a university can be positively enhanced if this institution appoints an accomplished scholar as president (vice chancellors). Furthermore, Zucker and Darby (2007, 2010) find that stars themselves, rather than the disembodied knowledge associated with them, are crucial for the entry of a broad range of high-tech startups.

1.4. Identification of Groups

Although there have been important advances in our understanding of the role research collaboration plays in the development of science, much remains yet to be explored. In particular, when characterizing and assessing research groups (collaborations) the Research Evaluation literature usually overlooks the endogenous characteristics of the research endeavor. In a typical assessment, the unit of analysis is defined ad hoc and is typically associated with an administrative boundary (i.e. by department or institution) rather than a true research group. One area that has provided much advance in the characterization of scientific collaboration and identifying communities in social networks is the social network literature. This body of work provides structure to research collaborations by developing community structure algorithms that identify groups within patterns of collaboration, in particular co-authorship networks (as well as

---

[4] Malmgren et al. (2010) define *mentorship fecundity* as the number of protégés a mentor trains.

in other types of ensembles). The idea is that nodes in a network represent authors of a paper and an edge establishes whether two authors have been co-authors in one or more papers. In addition, nodes within a community have dense connections with each other and sparse or null interactions with vertices outside their group.

Most algorithms operate by breaking a network into communities of nodes with dense connections within these groups and looser connections to other groups (Wasserman and Faust, 1994, p249-290; Girvan and Newman, 2002; Newman, 2004; Radicchi et al., 2004). For example, graph-partitioning algorithms try to break the network into some g number of groups with roughly the same node size (Kernighan, and Lin, 1970), while edge-removal algorithms identify the different communities by removing the links between groups with high edge centrality (Girvan and Newman, 2002)[5]. These methods have typically been developed to characterize generalizable patterns across a variety of social networks, from scientists to the production of Hollywood movies. They consider that networks are alike and the communities that arise are formed under the same mechanisms and following similar patterns. Therefore, by construction, they use the same algorithm, instead of exploring the peculiarities of how the science endeavor is organized. As a result, these views may fail to recognize and leverage specific characteristics that collaborative networks of researchers may have, which distinguish them from others.

As noted in the previous sections, another area where substantial work has been done in the last years is the one focused on quantifying the influence of eminent scientists on others. Initial research has often composed of case studies that recount the mentoring experience, cross

[5] Other popular alternatives are hierarchical clustering algorithms, which divide a network based on a hierarchy and a measure of similarity between pairs of vertices (Scott, 1988) and the clique percolation algorithm defines the different communities based on overlapping modular structures (Palla et al., 2005).

sectional studies or longitudinal analysis with usually short time frames, as well as small and often random data sets (e.g. Long et al., 1979; Reskin, 1979; Green and Bauer, 1995; Williamson and Cable, 2003; Judge et al., 2004; Paglis et al., 2006). More recently, there has been quantitative research in this domain, in particular looking at the impact of stars (Oettl, 2009; Azoulay et al., 2010; Waldinger, 2010). Yet, research has not really considered how theses stars interact with other scholars in the context of research teams, and their impact on the evolution of the system, especially in terms of researchers at the beginning of their career. Moreover, with the exception of Malmgren et al. (2010), research on mentorship has not looked at the extent to which protégés mimic their mentors' steps, performance and reputation.

Previous research on scientific cooperation has also typically considered collaboration in broad general terms, or focused on a particular research environment (like only collaborating with eminent scientists) failing to notice the combined effect different environments have on scientists, for example publishing with a highly productive researcher vs. a highly productive research group. Moreover, research has not inquired as to how these various setups affect productivity or impact.

Furthermore, with a few exceptions (Wagner et al., 2001; Gonzalez-Brambila and Veloso, 2007; Ordoñez-Matamoro et al., 2009; Horta et al., 2010), previous research in this area has mostly focused on the developed world. However, the ST&I community around the world has different characteristics (Nelson, 1993). This means that a better understanding of the factors that condition research output, impact and success in science requires an analysis of a diverse set of countries. In addition, ST&I systems have particular disparities between developing and developed nations. In developing nations there are fewer resources and infrastructure dedicated to research and development (R&D). Moreover, government funds most R&D and human as well as financial resources are centralized in a few institutions[6]. Thus, studying emerging economies provides a better understanding of the factors that influence the performance, impact and overall contribution of scientists in this environment (Nelson, 1993). In addition, studying this type of countries is relevant because these nations are actively developing and implementing policies to improve their S&T systems. Therefore, a better understanding of the factors that foster success at individual and aggregated levels could help leap forward their system.

This thesis begins by developing and testing a characterization and assessment method that recognizes the particular endogenous, or self-organizing characteristics of research groups. Instead of establishing an ad-hoc unit of analysis and assuming an unspecified network structure, the proposed method uses the notion that modern science is conducted primarily through a network of collaborators (or groups) who organize themselves around key researchers, often known as the principal investigators. Specifically, the method identifies Principal Investigators[7] (PIs), or key figures in an RG and then characterizes its boundaries using the pattern and strength of ties they have with their coauthors. In addition, it relies on the body of knowledge that each RG leverages, identified through the backward citations found in their published papers, to establish group 'knowledge footprints'. These footprints are used to evaluate the degree of structural similarity between groups, which is assessed through the degree of common cites to papers. Once all RG are characterized and their peers identified, one can measure and compare the performance/productivity of each RG.

---

[6] E.g. in Mexico in 2002 68% of the Gross Domestic Expenditure on Research and Development (GERD) was financed by the public sector (CONACYT, 2004, p16). In addition, in 2003 the National Autonomous University of Mexico (UNAM) had 27% of all the researchers belonging to the National Research System (SNI, http://www.conacyt.gob.mx/sni/) and received almost 50% of the federal R&D funding and four public institutions monopolized 92% of this budget (CONACYT, 2004, p24).

[7] For this work we define a *Principal Investigator* (PI) as an author with a high number of repeated connections, i.e. a researcher that has written several papers with a high number of coauthors.

This thesis continues by looking at the role that scientific stars (i.e. the most accomplished and salient researchers) have in a science system. In particular we assess how relevant these eminent scientist are for the development of a system. This means understanding how much they contribute to the output and impact of the system, as well as how influential are them in breeding the next generation of successful scientists, i.e. how successfully theirs protégés mimic their stellar performance. In addition it assesses how collaboration conditions the development of incoming scientists. In particular, we will look at the importance of the collaboration network of early co-authors for the productivity of new scientists and the likelihood that they also will become leading scientists.

Finally, following the previous perspective, but from a qualitative point of view, this thesis focuses on assessing the impact different forms of early collaborations have on the professional careers of new scientists. In particular it surveys a group of researchers (in an emerging economy) about their initial cooperation in science, the opportunities these early interactions opened to them and what they learned from these initial relationships.

Looking at these issues yields an important contribution to the literature on research collaboration, research evaluation and the science of science. There has been important progress in developing new approaches for assessing of research. There is also a growing knowledge based on methods that structure cooperation in science and identify its communities. There is a vast literature addressing the role research collaboration, and the collaborative environments that emerge from this interaction, plays in the development of science. But existing research on evaluation has hardly looked at how the endogenous characteristics of the research endeavor influence such assessments. By looking at the patterns and strength of collaboration to characterize teams and using their respective set of backward citations patterns (or knowledge footprints) to assess their similarity research administrators, scholars and policy makers can have better assessments of the research endeavor. In addition, current research on collaboration has focused on understanding the impact this phenomenon has throughout the professional career of a researcher or the influence a particular environment, like working with an eminent scientist, has on its development and performance. By looking at new scientists and the initial conditions they face (in terms of research milieus) we can better understand which will have a higher chance of succeeding in science and making the biggest contribution to the ST&I system.

To explore these issues, chapters 2 and 3 of this thesis use a database from Thomson Scientific[8] (Institute of Scientific Information, 2003) containing all papers published between 1980 and 2003 with at least one address in Mexico. This database contains the following information: article name, author(s), author(s), address(es), year of publication, journal, volume, pages, backward citations (i.e. references) and total number of citations received. From this database we selected all the papers published in Mexico in Physics and related areas (like applied physics, optics and material science among others) in the period of 1981-2003[9]. We chose these fields because they have been widely studied around the world (e.g. Collazo-Reyes et al. ,2004; Shrum et al., 2007) and Mexico has a long tradition of publishing in international peer reviewed journals, indexed by ISI in these areas (ISI, 2003; CONACYT, 2008). In addition, for chapter 3 we developed a questionnaire (appendix 1 has a sample of this instrument) that asks researchers about their initial interactions in science. The survey looks into the following topics: (1) how the researcher started his/her early collaborations in science, (2) what was the impact different environments and the interactions that happened within these settings had on incoming scientists,

---

[8] Formerly known as the Institute for Scientific Information (ISI)

[9] We only considered articles; this means that letters, notes and reviews were excluded. In addition, the extracted data have been undergone a detailed cleaning and then processed to bibliometric indicators.

(3) what opportunities these relationships opened, and (4) what the new researchers got out of these partnerships. We asked all the scientists that belong to the field of physics and related areas in Mexico that are part of CONACYT's National Research System[10] (SNI, acronym in Spanish) to participate in this survey.

1.5. Thesis structure

This thesis is organized into five chapters. Chapter one is this introduction. Chapter two (Research Groups Characterization and Assessment) draws from the literature of research evaluation and social network analysis to propose an evaluation method that takes into account the endogenous characteristics of research groups. Using data from the field of physics and related areas in Mexico, the chapter shows that the strength and frequency of the collaboration patterns are useful for identifying cohesive groups. In addition, this new technique allows scholars and policy makers to take into account the (expected) heterogeneity within institutions in their assessments. In addition, the overlap of common backward citation patterns and the benchmark at different levels of similarity in this cited work allows a departure from the established evaluation literature. This step allows potential evaluators to identify similar research groups, assess these groups and produce more meaningful comparisons and rankings. Furthermore, the research done by the different groups in Physics and related areas is (almost) non-redundant, i.e. the KFP overlap of these groups is relatively small, which means that each RG is (virtually) focused in one area of the research space.

---

[10] The SNI System was created 1980s by the Mexican Government to recognize the scientific and technological contribution of researchers in this country. The recognition is base on peer review evaluations and grants the appointment of National Researcher. This system has four levels, which are based on performance: candidate (which usually is the entry level to the system) and levels one, two and three; where the last level is the highest recognition within this system. Parallel to the appointment, the researcher receives an economic incentive based on the tier she belongs to (CONACYT, 2012)

Chapter 3 (Birth of prominent scientists) confirms our expectations and previous results that eminent scientist have a prime role in the development of a scientific system, especially within the context of an emerging economy like Mexico. In particular, in terms of productivity and visibility, this work shows that this elite group (defined as all scientists that are above the average productivity plus one standard deviation in our sample) published 42% of all publications and received 50% of all citations and bred 18% to 26% of new entrants. This work also shows that scientists that enter the system by the hand of a prominent researcher had higher productive and citation rates vis-à-vis scholars that did not publish their first manuscripts with an accomplished scientist. In addition, young researchers had an additional boost on their productivity if they had an early collaboration with a scientist that belonged to a highly productive research group, but these settings did not have any effect on the citation rate of new faculty. In terms of mimicking success, we find that scientists working at the beginning of their careers with eminent researchers tend to replicate the success of their mentors.

Chapter 4 (Learning and Opportunities in Collaborative Research Environments) suggests that, from the beginning of their career, star scientists have a different mindset with regards to collaboration when compared with non-star scientists. Starting scholars who are to become prominent researchers focused on acquiring new ideas and being exposed to the frontier of science through international collaboration and collaboration with star scientist, whereas less prominent ones are more about obtaining access to economic and physical resources (like specialized laboratories), and learning “basic” scientific skills (like publishing or research techniques).

Finally, Chapter 5 (Conclusions) discusses the importance of this work to current policy problems and highlights the most important contributions from this thesis. Included in the conclusions is a short discussion of future work that could help address the issues identified in this thesis.

Three appendices provide the full names and acronyms for the research organizations in chapter 2, the results of a sensitivity analysis to the regression models of chapter 3 and the survey used for chapter 4.

Chapter 2. Research Groups Characterization and Assessment

2.1. Introduction and Motivation

Throughout the last decades, tightening budgets and increasing competition, combined with a higher awareness on the outputs of science, have stimulated the development of new approaches towards the assessment of science (COSEPUP, 1999; Georghiou and Roessner, 2000, van Raan, 2000; Rip, 2000; Frederiksen, Hansson and Wenneberg, 2003). Current assessments have evolved from the classical peer review to an "informed" peer review, in which research is evaluated with the aid of quantitative benchmarks. Despite an important evolution, existing approaches towards scientific assessment still have a critical limitation: the boundaries of the unit of analysis are typically rigid (by individuals, institutes/departments, institutions, disciplines, regions, or countries), overlooking the unique and self-organizing characteristics of the research endeavor (Guimera et al., 2005). This means, for example, that present techniques have difficulty noting differences between low and top performing groups within a focal unit, say a university or even a department within a university. Likewise, benchmarking performance of university departments in a given area based solely on number of papers or citations fails to recognize that, for example, theoretical or experimental research profiles will necessarily imply different levels of publication and citation outputs. This renders a comparison based on average levels of productivity or impact for a broad area of limited value and potentially misleading. Furthermore, the performance of groups, or subunits (as described in Gläser, Spurling and Butler, 2004) cannot easily be measured. Finally, current methods are particularly limited when assessing interdisciplinary research groups (RGs) because it is difficult to ascribe these groups to a particular field of knowledge and measure their performance within this field in comparison with equivalent groups.

Along with the work on research evaluation, there has been important progress in the development of community structure algorithms that identify groups within a network. Most of these algorithms operate by dividing a network into communities of nodes with dense connections within these groups and looser connections to other groups (Wasserman and Faust, 1994, p249-290; Girvan and Newman, 2002; Newman, 2004; Radicchi et al., 2004). For example, graph-partitioning algorithms try to break the network into some g number of groups with roughly the same node size (Kernighan, and Lin, 1970), while edge-removal algorithms identify the different communities by removing the links between groups with high edge centrality (Girvan and Newman, 2002)[11]. These methods have typically been developed to characterize generalizable patterns across a variety of social networks, from scientists to the production of Hollywood movies. Therefore, by construction, they consider that networks are alike and the communities that arise are formed under the same mechanisms and follow similar patterns. Such approach may fail to recognize and leverage specific characteristics that collaborative networks of researchers may have, which distinguish them from others.

This research develops and tests a characterization and assessment method that recognizes the particular endogenous, or self-organizing characteristics of research groups. Instead of establishing an ad-hoc unit of analysis and assuming an unspecified network structure, the proposed method uses the notion that modern science is conducted primarily through a network of collaborators (or groups) who organize themselves around key researchers, often known as the principal investigators. Specifically, the method identifies Principal Investigators[12]

[11] Other popular alternatives are hierarchical clustering algorithms, which divide a network based on a hierarchy and a measure of similarity between pairs of vertices (Scott, 1988) and the clique percolation algorithm defines the different communities based on overlapping modular structures (Palla et al., 2005).

[12] For this work we define a *Principal Investigator* (PI) as an author with a high number of repeated connections, i.e. a researcher that has written several papers with a high number of coauthors.

(PIs), or key figures in an RG and then characterizes its boundaries using the pattern and strength of ties they have with their coauthors. In addition, it relies on the body of knowledge that each RG leverages, identified through the backward citations found in their published papers, to establish group ‘knowledge footprints’. These footprints are used to evaluate the degree of structural similarity between groups, which is assessed through the degree of common cites to papers. Once all RGs are characterized and their peers identified, one can measure and compare the performance/productivity of each RG.

The method is demonstrated by ranking research groups in Physics, Applied Physics/Condensed Matter/Materials Science and Optics in the leading institutions in Mexico, and showing how these groups are formed over time. This paper has two main results. First, it shows that the understanding of the scientific performance of an institution changes with a more careful account for the unit of analysis used in the assessment. Second, evaluations at the group level (using their knowledge footprint) provide more accurate assessments since they allow for appropriate comparisons within subfields of science. Third, this paper provides evidence that the performance of Mexican institutions in physics and related areas is highly heterogeneous within the institution itself. Finally, it shows that the research system is quite small, allowing less than two research groups (on average) in each subfield.

This paper is divided into five sections. First, it describes different types of evaluations and their limitations. Second, it introduces the theories that support our method, Bibliometric and Social Network Analysis, reviewing in particular the most common community structure algorithms. Third, it explains the method to identify the PIs of the networks and their groups, and assesses their absolute and relative performance. Fourth, the method is used to analyze the fields of Physics, Applied Physics and Optics in Mexico in the 1995-1999 period, showing in particular how to assess the absolute and relative performance of the groups in the country. Finally we present some policy implications.

2.2. Research Evaluation

In the last 30 years research evaluations and assessments have been on the rise. This trend has been fueled by budget stringencies, the need to better allocate scarce public resources and even a reassessment of the appropriate role of government in the economy (Papaconstantinou and Poltto, 1997, p. 9). In the specific context of Science and Technology (S&T), increasing costs and the desire to appropriately use the knowledge and results of these activities, combined with the need to further our understanding of the consequences of S&T policies, have spurred the use of a variety of assessment activities (Martin and Irvine, 1983; COSEPUP 1999; van Raan, 2000; Rip, 2000, Frederiksen, Hansson and Wenneberg 2003).

To accommodate these new S&T realities, several types of research assessments have been developed. Table 2.1 lists and describes the most common types of evaluations.

Table 2.1. Current Methods for Research Evaluation

Methods	Description
Bibliometric analysis	Assumes that publications, citations, and patent counts signal the work and productivity of a unit of analysis.
Economic rate of return	Used to estimate the economic benefits (such as rate of return) of research; gives a metric for the outcomes of research.
Peer Review	
Classic approach	Traditional method for evaluating science in which scientists continuously exercise self-evaluation and correction. Focuses on individual scientific products.
Modified approach	"Natural" development from the classical one, it incorporates issues that are not strictly cognitive. Focuses on group learning.
Case studies	"Historical accounts of the social and intellectual developments that led to key events in science or applications of science illuminate the discovery process in greater depth than other methods"
Retrospective analysis	Similar to case studies, but instead of focusing on one scientific or technological innovation it focuses on multiple cases.
Benchmarking	Used to assess whether a particular unit of analysis is at the cutting edge in terms of research, education or other measures.

Source: COSEPUP (1999, p 18-22); van Raan (2000); and Frederiksen, Hansson and Wenneberg (2003).

Each method is substantially useful in its own way, but also has significant drawbacks (Table 2.2). For all established methods noted above, important additional limitations exist. One that stands out in particular is the fact that the boundaries of the unit of analysis are defined ad-hoc, overlooking the endogenous characteristics of the research groups that are formed within such units. Furthermore, these assessments often assign broad cohort groups and often consider all units within a field as if they were homogenous in terms of the knowledge they use and their relative output and impact. The method proposed in this study aims to address these limitations.

Table 2.2. Pros and cons in current evaluation methods

Methods	Pro	Con
Bibliometric analysis	Quantitative; useful on aggregate basis to evaluate quality for some programs and fields	At best, measures only quantity; not useful across all programs & fields; comparisons across fields or countries difficult; can be artificially influenced
Economic rate of return	Quantitative; shows economic benefits of research	Measures only financial benefits, not social benefits; time separating research from economic benefit is often long; not useful across all programs and fields
Peer Review	Well-understood method and practices; provides evaluation of quality of research and sometimes other factors	Focuses primarily on research quality; other elements are secondary; evaluation usually of research projects, not programs; great variance across agencies; concerns regarding use of "old boy network"; results depend on involvement of high-quality people in process
Case studies	Provides understanding of effects of institutional, organizational, and technical factors influencing research process, so process can be improved; illustrates all types of benefits of research process	Happenstance cases not comparable across programs; focus on cases that might involve many programs or fields making it difficult to assess federal-program benefit
Retrospective analysis	Useful for identifying linkages between federal programs and innovations over long intervals of research investment	Not useful as a short-term evaluation tool because of long interval between research and practical outcomes
Benchmarking	Provides a tool for comparison across programs and countries	Focused on fields, not federal research programs

Source: COSEPUP (1999, p 18-22).

2.3. Theoretical Background

The proposed method is based on the notion that Research Groups (RG) are at the core of the research endeavor and that assessment instruments should have the RG at the heart of the analysis. The development of our method has three main steps. First, we use bibliometric analysis to chart the co-authorship network within a field and define the patterns of collaboration among this set of connections. Second, we use social network analysis and the institutional affiliation of the individuals and groups to delimit the boundaries of an RG. In particular, we draw on the notion of lambda sets (LSs) and cliques; we use the output of the LSs algorithm to find the key people, which we identify as PIs in the network and cliques to identify the cohesive group of direct collaborators of these researchers and thus establish the boundaries of the group. Third, we again rely in bibliometric analysis to establish the level of similarity between RGs. In particular, we use co-citation across groups to establish a metric of knowledge distance and identify peer groups. Finally, we assess group performance with respect to publication output and citation counts (normalized by group size), comparing groups against their peers.

2.3.1. Co-authorship as a form of collaboration and network topology

For this study we follow the common practice in the field of using multiple-author papers as an indicator of collaborative activity within a field (Newman, 2001; Melin and Persson, 1996; Katz and Martin, 1997; Newman, 2004b; Wagner and Leydesdorff, 2005). The idea is that co-authorship is one of the most tangible and well-documented forms of scientific collaboration and the output of these interactions creates a 'co-authorship network' (Newman, 2001; Glänzel and Schubert, 2004) like the one depicted in figure 2.1. In this type of networks, a node represents an author of a paper and an edge establishes whether two authors have been coauthors in one or more papers. For this work, we use the number of co-authored papers (between authors) as an indication of the strength of the collaboration, i.e. higher numbers of papers represent stronger levels of collaboration.

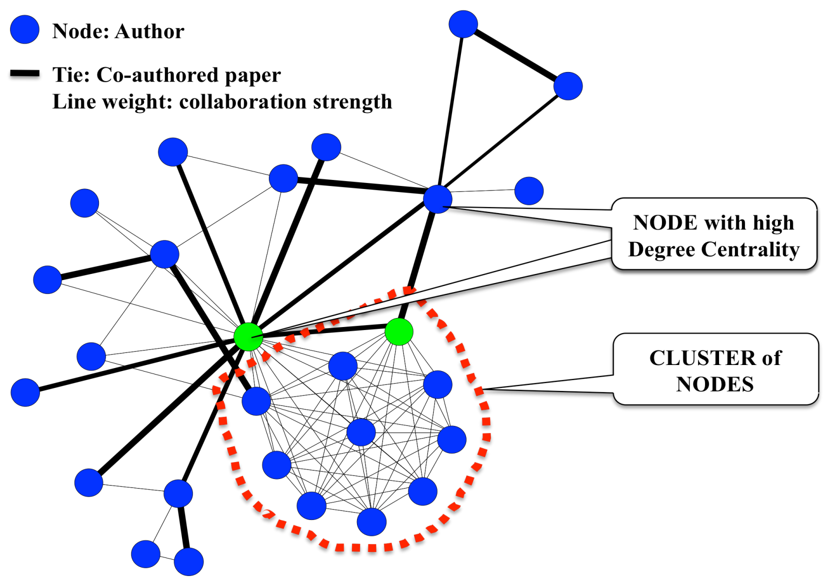


Figure 2.1. Co-authorship network. This network provides a hypothetical example of a co-authorship network, where the nodes represent the authors and the ties if two authors have written one or more papers together. In addition, the strength of the collaboration (weights of the line) is based on the number of co-authored papers.

2.3.2. Network Analysis

In its simplest form, a (social) network is a diagram of all of the relevant links between a certain group of nodes (or actors) that provides a means to visualize existing and potential interactions. Within the study of science and technology, Rogers, Bozeman and Chompalov (2001) state that networks can be used as guiding metaphors and as techniques "to measure structural properties of the ensemble". These authors classify these types of studies by "the level of analysis that is given by the nature of the actors that will be placed at the nodes" (i.e. nodes can be individuals, teams, departments or institutions), by the nature of the links between nodes (interaction networks vs. position networks)[13], and by the domain in which the actors belong (intra-organizational vs. inter-organizational)[14].

In this work, we focus our analysis on the ties that emerge through co-authorship, which represent “formal” social interactions among researchers, and define an RG based on its levels of cohesiveness and the connectivity between co-authors. A network analysis of these co-authorship links is particularly useful because it will allow us to explore the structural configuration of social interactions, which provides a powerful model for the underlying social arrangement (Scott, 1988) that we are interested in understanding.

Identification of groups and communities within networks

The rich set of connections between nodes in many networks produce heterogeneous structures where cohesive subgroups of nodes will have relatively strong, direct, intense, frequent, or positive ties within a subgroup but fewer edges between subgroups Wasserman and Faust (1994, p. 249); creating a structure like the one depicted in figure 2.2.

---

[13] In **Interaction networks**, the links represent actual information exchanges or other communication events between actors; whereas in **position networks**, they represent relationships established by the relative positions of the actors in the system (Rogers, Bozeman and Chompalov, 2001)

[14] **Intra-organizational** studies only consider links between actors inside the boundaries of a single organization; in contrast, **inter-organizational** studies consider links between actors across organizational boundaries (Rogers, Bozeman and Chompalov, 2001)

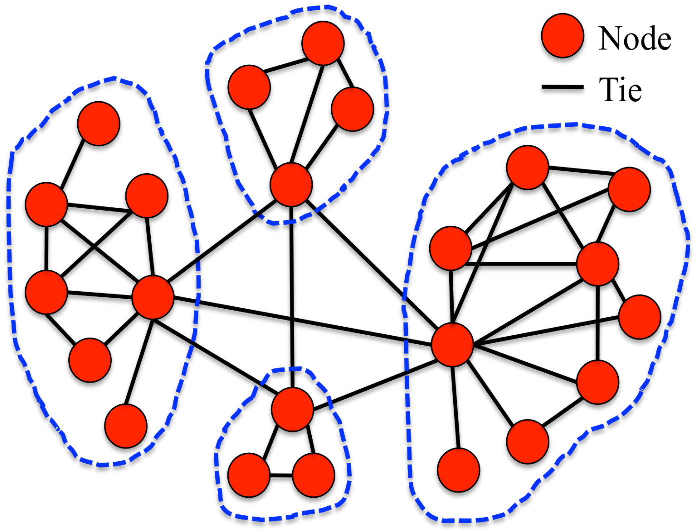


Figure 2.2. Subgroup/community structure within a network.

The question of how to identify these subgroups within a network has been addressed before in a variety of contexts. From a theoretical point of view, Wasserman and Faust (1994, p. 251) uncover four mechanisms by which these clusters of nodes can be formed, and based on these mechanisms they identify eight methods for finding this type of ensembles within any type of networks: cliques, n-cliques, n-clans, n-clubs, k-plexes, k-cores, LS Sets, Lambda Sets (Wasserman and Faust, 1994, p. 249-290). Table 2.3 summarizes these methods.

Table 2.3. Methods used to delimit subgroup

Mechanism	Method	Definition
Mutuality of ties	cliques	A clique is a maximal complete subgraph with at least three nodes. It is a subset of nodes, all of which are adjacent to each other, and there are no other nodes that are also adjacent to all of the members of a clique.
Reachability and diameter (or closeness)	n-cliques	An n-clique is a maximal subgraph in which the largest geodesic distance between any two nodes is no greater than n. When n = 1, the subgraphs are cliques.
	n-clans	An n-clan is an n-clique in which the geodesic distance, d(i,j), between all nodes in the subgraph is no greater than n for paths within the subgraph.
	n-clubs	An n-club is a subgraph in which the distance between all nodes within the subgraph is less than or equal to n; furthermore, no nodes can be added that also have geodesic distance n or less from all members of the subgraph.
Nodal Degree (frequency of ties members)	k-plexes	A k-plexes is a maximal subgraph in which each node may be lacking ties no more than k subgraph members. When k = 1, the subgraph is a clique.
	k-cores	A k-cores is a subgraph in which each node is adjacent to at least a minimum number, k, of the other nodes in the subgraph.
Frequency of ties within vs. outside subgroups	LS Sets	An LS set is a subgroup definition that compares ties within the subgroup to ties outside the subgroup by focusing on the greater frequencies of ties among subgroup members compared to the ties between subgroup members to outsiders.
	Lambda Sets	A lambda set is a cohesive subset that is relatively robust in terms of its connectivity, i.e. it is difficult to disconnect by the removal of lines from the subgraph. Sub groups are defined by varying the number of ties, d, within clusters of nodes; the more ties you have to drop (within a group) the more cohesive this group is.

Source: Based on Wasserman and Faust (1994, pp. 251-267).

From a large-scale/real-world network perspective, several algorithms have been developed for the purpose of identifying communities (or groups) within this type of networks, like work assignment in parallel computing, co-authorship networks, Protein networks, and social networks (Kernighan and Lin, 1970; Ravasz et al., 2002; Spirin and Mirny, 2003;Newman 2004a; Newman 2004c; Radicci et al., 2004; Palla et al., 2005). These algorithms typically perform such identification by assuming a general structure for the network and focusing on how to divide it into a set of disjoint communities. Two main approaches have been followed [Newman, 2004a]: the computer science approach, where the nodes of a network are divided into sets of roughly equal size, and sociological approach, where groups try to mimic reality and groups differ in size.

Early algorithms for partitioning a network attempted to find the minimum-cut to partition the network into two groups with roughly the same number of nodes (Ford and Fulkerson, 1962; Kernighan and Lin, 1970). Although useful for some applications in parallel computing, these algorithms are limited because they only split the graph into equivalent parts, and overlook the underlying structure of the communities. The hierarchical clustering method goes beyond the minimum-cut methods and can be customized to fit the underlying social structure of a network by varying the strength of connection between vertices, i.e. adding or removing edges to or from the network. (Scott, 2000). Still, this method assumes that such a measure exists for a given network, and that it is the only way in which vertices are categorized into groups. In addition, this method does not produce a definite number of groups, relying on some other complementary decision mechanism to establish a final network partition (Newman. 2004a).

More recent approaches overcome most limitations of the algorithms described above by iteratively removing edges and measuring the strength of the community structure found by the algorithm at each iteration. Yet, they still rely on the assumption that the community structure is generally consistent across any type of real world networks. For example, Newman's Modularity Maximizing algorithm identifies groups based on the modularity metric, which is calculated using the expected number of edges between members of a group (Newman, 2006). Likewise, the Girvan-Newman algorithm forms groups based on the betweenness of edges[15], (Newman,

[15] *Edge betweenness* is a measure that favors edges that lie between communities and disfavors those that lie inside communities. The betweenness of an edge is the number of shortest paths between all pairs of nodes in a network that run through it. When a network is made of tightly bound groups, loosely interconnected, all shortest paths between vertices in different groups have to go through the few intergroup connections, which therefore have a large betweenness value (Radicci et al, 2004).

2004). Table 2.4 summarizes the main algorithms used to partition a network and their limitations.

Table 2.4. Algorithms used to partition a network

Algorithm	Basic Measures	Limitations
Ford-Fulkerson Kerighan-Lin	Maximum-Flow/Minimum-Cut	• Only splits into two groups • Lacks consideration of underlying structure
Hierarchical Clustering	Similarity function	• Does not output definite set of groups. • Limited to a notion of similarity • Lacks consideration of underlying structure.
Modularity Maximizing	Modularity	• Lacks consideration of underlying structure.
Girvan-Newman	Modularity, Edge Betweenness	• Lacks consideration of underlying structure

2.3.3. Bibliometric Analysis

Since the 1970s performance analysis, typically based on publication output and received citation (Martin and Irvine, 1983; van Raan, 2000; Kane, 2001; van Raan, 2005), has been widely used in evaluative bibliometrics (see Noyons, Moed and Luwel (1999) and van Raan, (2004) for a historical development). This method is based on the premise that an article will be "published in a referred journal only when expert reviewers and the editor approve its quality" and it will be "cited by other researchers as recognition of its authority" (COSEPUP, 1999, p. 18; van Raan, 2004; Thomson-ISI, 2003)

Another important dimension in bibliometric analysis is science mapping, which typically relies on co-citation analysis (Small, 1973; Narin, 1976; Narin, 1978; Small, 1978; Leydesdorff, 1987; Gmür, 2003). Co-citation analysis looks at the structure and development of scientific communities and areas. This methodology is based on the notion that a citation is a valid and reliable indicator of scientific communication, and this measure signals the relevance of an article (Small, 1978; Garfield, 1979; Gmür, 2003) or a scientist. The specialized literature identifies a co-citation if two publications or authors cite the same reference and uses the number of co-citations as a measure of similarity of content or proximity between the two publications or authors (Gmür, 2003). In the last 30 years, two approaches have been developed within this framework, namely document co-citation (focused on documents and publications with peer-review procedures) and author co-citation (focused on researchers). In previous work, document co-citation (Small, 1973; Small, 1977) and author co-citation analysis (White and McCain, 1998), as well as co-word analysis[16] (Callon et al., 1983; Noyons, 2004), have been used to map the structure of scientific and technological fields and their evolution over time. These studies have aimed at "identifying and analyzing emerging research specialties or 'hot' topics of great strategic or technological importance, their actors, and their relationships to other areas of research" (Moed, 2005, p.17). For this paper, document co-citation analysis is useful because it will allow us to measure how close (or far apart) groups are in terms of knowledge.

2.4. Method

In this section we discuss the method developed for the characterization and assessment of research groups. This method has six steps. First, the collaboration patterns among researchers will be characterized by mapping a co-authorship network within a certain period of time and area(s) of knowledge. Second, the most salient researchers of the co-authorship network will be identified. These will be considered the principal investigators (PIs) and all the collaborators will be attached to these key people. Third, different research groups (RGs) will be defined based on the PIs of the network, their collaborators and the patterns of links among these collaborators and between these and the PI. Fourth, the knowledge base that each group relies on, which we will

---

[16] "Co-word analysis deals directly with sets of terms shared by documents instead of with shared citations. Therefore, it maps the pertinent literature directly from the interactions of key terms instead of from the interactions of citations" (Coulter et al., 1998)

term Knowledge Footprint[17] (KFP) is delimited and a co-citation analysis is performed with these footprints to establish the distance between groups (i.e. overlap of KFP between groups). Fifth, the performance of each RG is measured using (normalized) total scientific output and citation counts. Finally, the performance of each RG is compared against its peers. Figure 2.3 shows a conceptual representation of the method.

In the remainder of this section we will explain in detail each step and illustrate the application of the method by relying on the co-authorship network centered on the Physicist Jerzy Plebanski[18] and all the researchers that have a path length[19] of three or less from him (i.e. his coauthors, the coauthors of his coauthors and the coauthors of the coauthors of his coauthors) for all the papers published by these scientists within the 1997 and 1999 with at least one institution in Mexico.

---

[17] The *knowledge footprint* (KFP) for group *i* is the union of all the backward citations used by all members of a group in all of their papers within a specific time frame.
[18] We chose Jerzy Plebanski (1928-2005) to exemplify the method because he is a well known Polish Physicist that worked in Mexico for several years (CINVESTAV, 2008)
[19] The length of the path is the number of lines in a path (Wasserman and Faust; 1994, p. 107). "A path is a walk in which all the nodes and all the lines are distinct (Wasserman and Faust; *ibid*.)

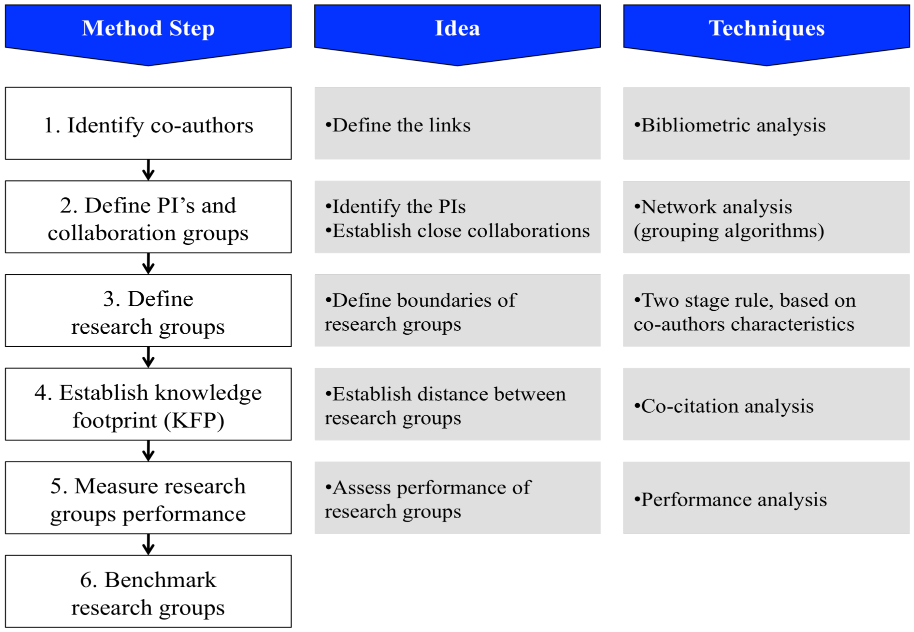


Figure 2.3. Method steps. This figure shows the steps of the method for the characterization, assessment and benchmark of research groups.

2.4.1. Network Definition

In this step, we define the patterns of collaboration within a field of knowledge by looking at a specific co-authorship network, within a certain period of time. As stated previously, a node will represent a researcher, a link will denote a co-authored paper and a weight of the link will capture the number of co-authored papers. This step will produce a weighted NxN adjacency matrix (author by author) where the w_{ij} values of each cell indicate the strength or frequency of the relation (co-authorship) between authors i and j. Figure 2.4 presents Plebanski's network and all the researchers that have a path length of three or less from him; in this graph the gray squares represent the authors and the links between them represent their common publications. The other elements in this figure are explained in the next section.

2.4.2. Identifying the Collaboration Groups

In this step all the collaboration groups (CGs) are demarcated. For the purpose of the study, a CG will consist of a PI and all its co-authors that have published a paper with her and at least another co-author of the PI; dyads of the PI (i.e. individual authors that have ONLY co-authored papers with the PI and with no other scientist in the PI's collaboration group) will be excluded from this group. We impose this restriction because ensembles of three or more nodes produce relations/groups that have higher quality, and are more dynamic and stable than dyads (Krackhardt, 1999). Moreover, it is rather intuitive to consider that a research group requires at least 3 participants.

This procedure entails three stages. First, the key scientists, or Principal Investigators (PIs), within a co-authorship network are identified by using the concept of lambda set (LS) and calculating the lambda set level (LSL) for each author. We use LS because this algorithm calculates the maximum number of ties that need to be removed so a node becomes completely isolated from the network. This means that by dropping and counting the ties at the node level this procedure gives a measure of the importance of each researcher, the LSL[20]. This stage produces a list of researchers decreasing in order of importance; measure in the number of ties that need to be removed so a researcher becomes completely isolated. Moreover, all the direct collaborators that have published a paper with the PI and at least an additional coauthor are identified using the concept of clique; this is done because this algorithm allows us to identify all

[20] The LSL L_R for a researcher R is defined as maximum flow f_{RT} between R and T where f_{RT} is greater than the maximum flow f_{RU} between R and any other researcher $\underline{U}$. By the max-flow min-cut theorem, this is essentially a measure of the maximum number of edges that need to be cut (or maximum number papers removed from the network), such that R is no longer connected to some other researcher (Ford and Fulkerson, 1962).
The max-flow min-cut theorem states that the size of the maximum flow, or the total amount of flow that can exist between source node s and target node t using the edges connecting s and t, is equal to the size of the minimum-cut between s and t.

the nodes that are adjacent to a node (in this case the PI) in groups of three or more nodes (Wasserman and Faust; 1994, pp. 254). This stage creates a set of cliques centered at the PI. Finally, for each PI we define its collaboration group as the union of all its cliques.

This step will produce a set of collaboration groups (centered at the PIs of the network) that might share one or more scientist on the interface of these groups; the challenge here is then to assign these scientists to only one group and create a set of disjointed groups. Figure 2.4 shows where Plebanski stands in this network, as well the other key neighbor scientists highlighted in color circles. In addition, this figure depicts (for illustration purposes) all the direct collaborators of Plebanski and Matos (squares and pentagons), and defines their respective collaboration groups (in broken lines). For this example we can see that “Obregon O” belong at least to two CGs, one centered at “Plebanski JF” and the other to “Matos T”.

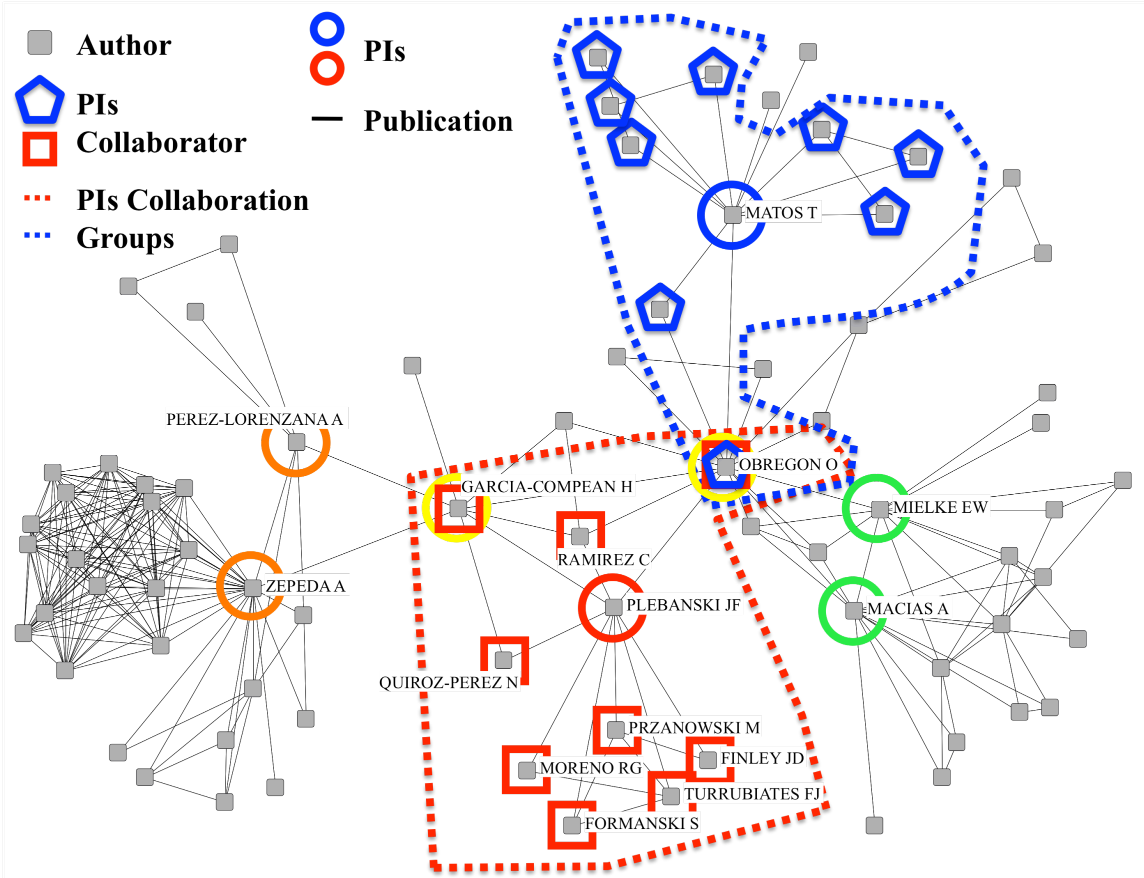


Figure 2.4. Identifying the PIs of the network. This figure shows for the Plebanski Network the different PIs (Plebanski, Matos, Perez-Lorenzana, Zapeda, Garcia-Compean, Obregon, Mielke and Macias) of this ensemble. In addition, it shows the Collaboration groups (CGs) of Plebanski and Matos and the interface of these CGs (Obregon).

2.4.3. Research Group Delimitation

In this step we define the boundaries of the research groups (RG) by using a two-stage rule on the scientists that are at the interface of two or more CGs. This means that, for each of these scientists, we first compare the percentage of co-authors she has in each group and assign her to the group where this percentage is higher. If for some reason this percentage is the same between two or more CGs we use the institutional profile[21] of the researcher and each CG. We

[21] The institutional profile of an author (or CG) is defined as a vector that contains in each cell the number of papers an author has published in a specific institution divided by the total number of papers this person has published, this

assign this researcher to the group that has the highest similarity between institutional profiles. We iterate this step until all researchers are assigned to one and only one RG. Figure 2.5 shows the four researchers within Plebanski's CG that are also part of other CGs and highlights all the CGs for O. Obregon, one of the scientist in the interface of Plebanski's CG.

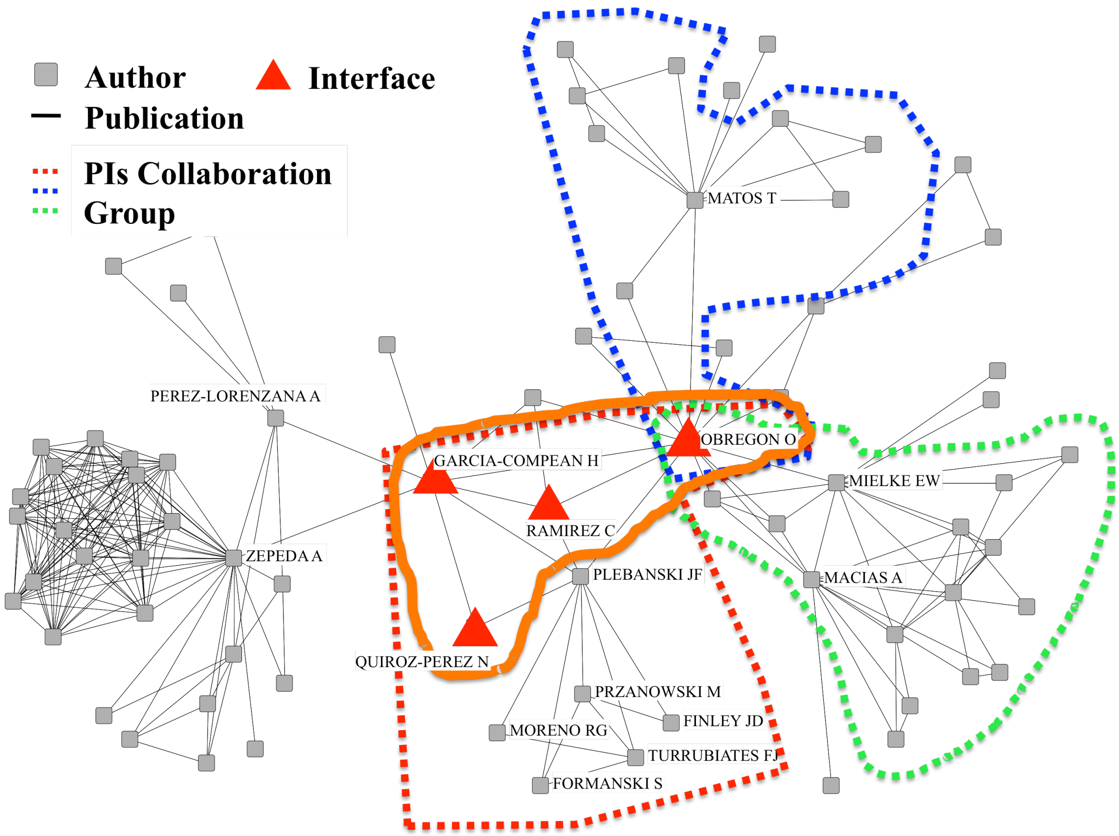


Figure 2.5. Interface for J.F. Plebanski's collaboration group.

After taking these steps, the universe of analysis will be clustered in a set of RG, as well as various researchers that are not integrated in any research group, as defined by the method.

---

means that if we have four institutions, A, B, C, D; and an author has published 5 papers in institution A, 3 in institution B, 2 in institution C and 0 in institution D her institutional profile is (0.5, 0.3, 0.2, 0.0). This concept can be extended to the collaboration groups by counting for each institution the number of paper each author of the CG has published in this institutions and dividing this number by the total number of papers the CG has.

Figure 2.6 shows how Plebanski's network breaks down into research groups after application of the algorithm.

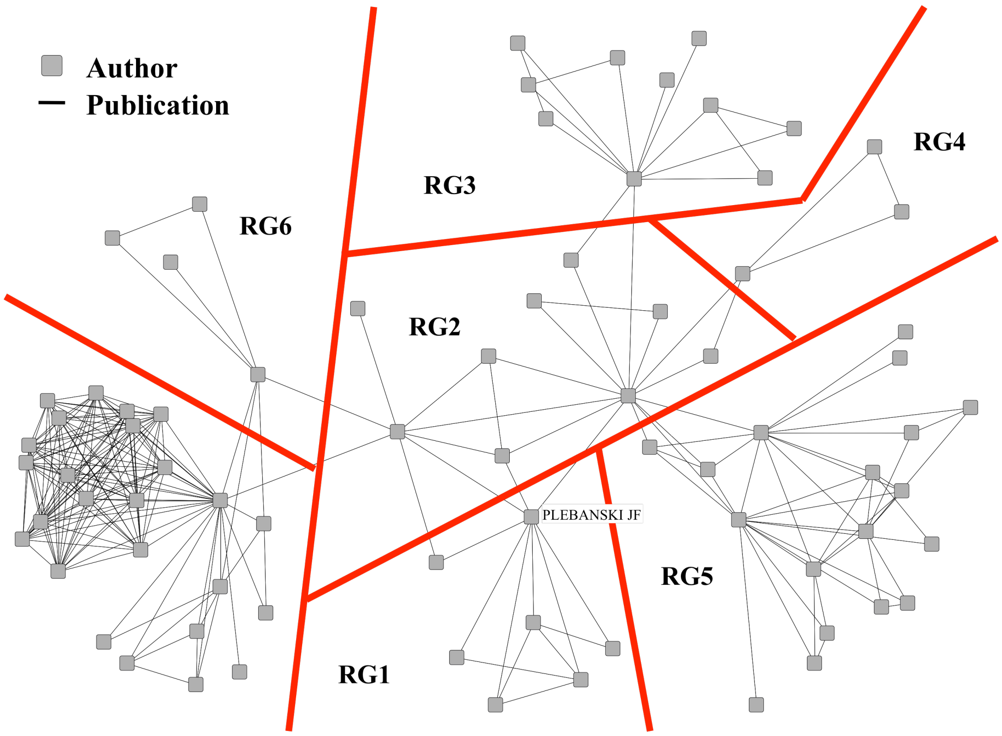


Figure 2.6. Plebanski's three degree co-authorship network.

2.4.4. Knowledge footprint and group similarity

In this step of the method we define the level of similarity (or distance) between groups by identifying what we will define as the Knowledge Footprint of each group and then establishing the overlap between footprints. For such purpose, we characterize the Knowledge Footprint (KFP) of an RG as the union of all backward citations found in the work published by each group, and define the level of similarity between groups by performing a co-citation analysis between the footprints of the groups. A group is said to be completely similar to another if the KFP of the former entirely overlaps the KFP of the latter, in which case we say that the level of similarity is 100%. In contrast, a group is dissimilar to another group if the intersection of their citations is disjointed; in this case we say that the level of similarity is 0%.

The level of similarity is used to identify benchmark groups within a research system. We will use a one, five, ten, twenty-five and fifty percent overlap in KFP as baseline levels for similarity between groups. Figure 2.7 shows the KFP for the Pelbanski network. In this case the yellow heptagon is Pelbanski's RG, the blue ones are the other RGs, the KFP overlap (links between triangles) and the different levels of similarities (line weight). Furthermore, this figure reveals the degree of structural similarity between groups by varying the level of similarity, from 1% to 25%.

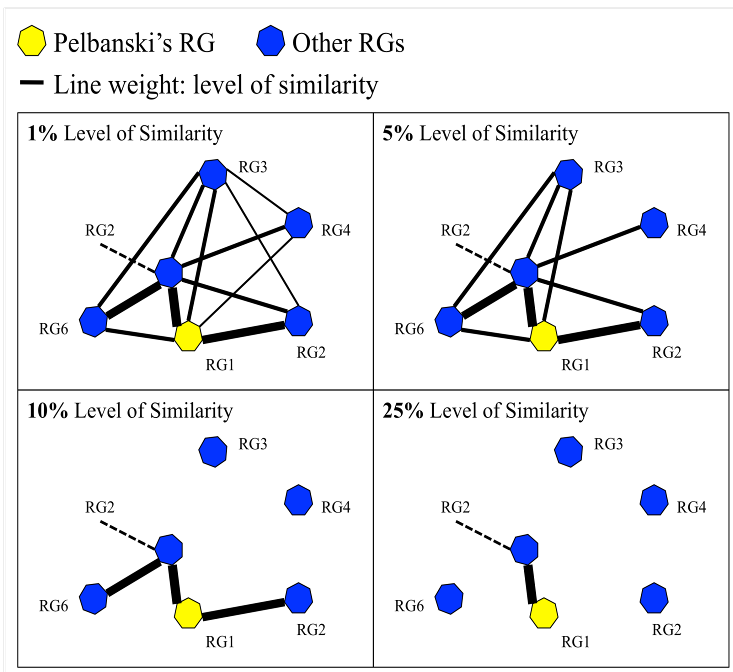


Figure 2.7. Knowledge Footprint Hypothetical Example.

2.4.5. Scientific output and performance

Once we have identified the relevant research groups and their benchmark peers, we need to compare their performance. To measure the scientific performance of each group, we use total scientific output (papers published) divided by the number of researchers in a group as a proxy for a normalized measure of scientific productivity, and citation counts by group size as a proxy for a normalized measure of scientific impact (formula below).

$$\text{Normalized total scientific output (citation count)} = \frac{\sum \text{Publications (citations) per group}}{\sum \text{Reserchers per group}} \qquad (1)$$

2.5. Demonstration of the Method

In this section we first describe the data used to demonstrate the application of the method described in the previous section. Then we demonstrate how this method can be used to identify groups and assess their performance.

2.5.1. Data

To test the proposed method, we use a database from Thomson Scientific[22] (2004), formerly known as the Institute for Scientific Information (ISI), owned by the Mexican Council for Science and Technology (CONACYT). This database contains all papers published between 1980 and 2003 with at least one address in Mexico[23]. This database contains the following

[22] Thomson Scientific became in 2008 the Scientific business unit Thomson Reuters after the merger of the Thomson Corporation with Reuters in that year.
[23] In the last stage of our analysis we realized that this database also contained publications with at least one address in New Mexico and none in Mexico. We think that Thomson Scientific might have created this database with the key word *Mexico*, including all the papers from *Mexico* and *New Mexico*. We preserved the papers with all the addresses because we did not want to mistakenly eliminate useful data, however our primary concern was to concentrate on Mexico.

information: article name, author(s), author(s), address(es), year of publication, journal, volume, pages, backward citations (i.e. references) and total number of citations received.

To illustrate the application of the method, we selected all the papers published in Mexico in the areas of Physics, Applied Physics/Condensed Matter/Materials Science and Optics in the period of 1990-1999. We chose these areas because Physics and its related areas have been widely studied around the world (e.g. Braun et al., 1992; Nagpaul and Sharma, 1994; Miquel et al., 1995; Marx and Hoffmann, 2011). Moreover, Mexico has a long tradition of publishing in international peer reviewed journals indexed by ISI in these areas (Narvaez-Berthelemot et al., 1992; Russell, 1995; Gómez et al., 1999; Collazo-Reyes and Luna-Morales, 2002; Collazo-Reyes et al., 2004; Russell et al., 2007). The period for the analysis considers the most recent publications available in the database, while allowing for a citation window of five years. (i.e. for the papers published in 1990 we restricted the citation count to the period 1990-1994, for 1991 we chose a 1991-1995 window, and so on). For the analysis, we used five-year periods, starting with 1990-1994 and ending with 1995-1999.

2.5.2. Main outcomes

Institutional Benchmarks

Table 2.5 shows a traditional assessment done at the level of the institution (University or Research Center). From this table, it can be seen that the Autonomous National University of Mexico (UNAM) is the leading institution in absolute terms (for publications and citations) and ranks in the top ten if both measures are normalized by the total number of researchers ascribed to this institution.

Table 2.5. Publications and Citations in Physics and related areas for Selected Institutions, 1995-1999

Institution*	Articles Total (Rank)	Articles Per researcher (Rank)	Citations Total (Rank)	Citations Per researcher (Rank)	Researchers Total
UNAM	1376 (1)	1.54 (8)	5892 (1)	6.61 (7)	891
CINVESTAV	507 (2)	1.47 (9)	1957 (2)	5.70 (11)	343
UAM I	307 (3)	1.64 (6)	1211 (3)	6.47 (8)	187
BUAP	256 (4)	1.44 (10)	787 (6)	4.44 (15)	177
CIO	210 (5)	1.89 (4)	1079 (4)	9.72 (2)	111
INAOE	170 (6)	1.61 (7)	832 (5)	7.92 (3)	105
IPN	164 (7)	1.15 (19)	364 (11)	2.56 (23)	142
UASLP	136 (8)	1.43 (11)	660 (8)	6.94 (6)	95
CICESE	123 (9)	1.21 (17)	719 (7)	7.11 (5)	101
UniGuan	123 (9)	1.89 (3)	467 (9)	7.18 (4)	65
UniSon	112 (10)	1.36 (13)	392 (10)	4.78 (13)	82
ININ	90 (11)	1.05 (22)	282 (12)	3.31 (17)	85
UAM A	61 (12)	1.35 (14)	137 (16)	3.04 (20)	45
UAEdoMor	60 (13)	1.42 (12)	243 (14)	5.78 (10)	42
UniGDL	50 (14)	2.08 (2)	113 (18)	4.70 (14)	24
UAZ	42 (15)	1.82 (5)	138 (15)	6 (9)	23
UAQ	34 (16)	0.91 (25)	120 (17)	3.24 (19)	37
UniMich	31 (17)	0.96 (23)	50 (22)	1.56 (24)	32
IMP	26 (18)	0.96 (24)	81 (19)	3 (21)	27
ITESM	26 (18)	1.3 (15)	66 (20)	3.3 (18)	20
CIMAT	23 (19)	2.3 (1)	267 (13)	26.7 (1)	10
UDLAP	17 (20)	1.13 (20)	18 (24)	1.2 (25)	15
UAEdoMex	17 (20)	1.21 (18)	36 (23)	2.57 (22)	14
UANL	15 (21)	1.07 (21)	51 (21)	3.64 (16)	14
UABC	11 (22)	1.22 (16)	50 (22)	5.55 (12)	9

* For a full description of the institutional acronyms see table A1 in the appendix.

In Table 2.6, six institutions have been broken down according to administrative boundaries[24]. These depend on the nature of the institution and may reflect, for example, a research center focused just on research; an institute or department training graduate students and doing research; or a school primarily training undergrad students and doing little research. This Table reflects the typical level of detail allowed by existing methods. It already provides a more complete perspective than a high level institutional characterization. For example, the Physics Institute in UNAM (INST FIS) has a stellar performance (in total output in the area of physics),

[24] Only UNAM, CINVESTAV, UAM-I, BUAP, IPN and USLAP were broken down by department because ISI did not record the department for the remaining ones.

stronger than the Faculty of Science (FAC CIENCIAS) and much different from the Applied Math and Systems Research Institute (IIMAS). By contrast, in BUAP (fourth place at the institutional level) the Physics Institute (IFBUAP) performs above eight departments of CINVESTAV (second place at the institutional level).

However, if we would now look at the patterns of co-authorship and collaboration and identify research groups based on these patterns, a different and much more complete perspective emerges. Table 2.7 shows the number of groups[25] identified along two dimensions: the number of institutions[26] that are represented in the group (single vs. multiple institutions) and whether they belong to the top 25 Mexican institutions. The table shows that the method can identify 297 research groups. However, only 191 are relevant units of analysis (i.e. the group is conformed by at least three researchers from one of the top 25 Mexican institutions). Table 2.8 provides the summary statistics for these groups and table 2.9 gives the summary statistics for 19 of the top 25 Mexican institutions; these numbers differ because six institutions did not have sufficient critical mass to have groups with three researchers at the same institution.

---

[25] For this analysis a group should have at least three researchers and published two or more papers within a certain period.

[26] Single institution groups (with three or more researchers in a particular institution) are used to properly assess the performance of this institution. Multiple institutions appear as a baseline application of the method because the initial criterion is only co-authorship in the context of a clique. Only when defining the boundary of the groups taking in consideration the institution can we get to groups that are meaningful from the point of view of benchmarking.

Table 2.6. Publications and Citations in Physics and related areas for Selected Institutions, 1995-1999

	Articles		Citations		Researchers
	Total	Per researcher	Total	Per researcher	Total
Institution*	(Rank)	(Rank)	(Rank)	(Rank)	
UNAM					
UNAM - INST FIS	566 (1)	1.60 (16)	2878 (1)	8.17 (8)	352
UNAM - INST INVEST MAT	284 (3)	1.37 (26)	1104 (3)	5.35 (24)	206
UNAM - INST CIENCIAS NUCL	231 (4)	1.90 (5)	1039 (5)	8.58 (7)	121
UNAM - FAC CIENCIAS	84 (13)	1.4 (24)	172 (21)	2.86 (45)	60
UNAM - CTR INVEST ENERGIA	61 (17)	1.24 (33)	226 (19)	4.61 (31)	49
UNAM - CTR INSTRUMENTOS	52 (19)	1.44 (21)	129 (26)	3.58 (36)	36
UNAM - CTR CIENCIA MAT CONDENSADA	43 (21)	1.48 (19)	102 (33)	3.51 (37)	29
UNAM - INST QUIM	32 (25)	2.28 (3)	111 (29)	7.92 (9)	14
UNAM - FAC QUIM	28 (28)	1.12 (44)	105 (32)	4.2 (33)	25
UNAM - IIMAS	23 (31)	1.53 (18)	63 (40)	4.2 (33)	15
UNAM - INST MATEMAT	22 (32)	1.04 (48)	107 (31)	5.09 (27)	21
UNAM - INST ASTRON	17 (34)	1.41 (23)	134 (25)	11.1 (2)	12
UNAM - INST GEOFIS	13 (38)	1.3 (31)	14 (51)	1.4 (55)	10
UNAM - CTR INT CIENCIAS	13 (38)	1.08 (45)	108 (30)	9 (6)	12
UNAM - CTR CIENCIAS FIS	7 (42)	0.77 (57)	66 (39)	7.33 (11)	9
CINVESTAV					
CINVESTAV - DEPT FIS	357 (2)	1.73 (11)	1410 (2)	6.84 (16)	206
CINVESTAV - DEPT FIS APLICADA	60 (18)	1.39 (25)	298 (13)	6.93 (15)	43
CINVESTAV - DEPT INGN ELECT	40 (23)	1.17 (39)	95 (35)	2.79 (47)	34
CINVESTAV - DEPT QUIM	11 (39)	0.91 (53)	36 (45)	3 (44)	12
CINVESTAV - DEPT MATEMAT	8 (41)	1.14 (42)	33 (47)	4.71 (29)	7
UAM I					
UAM I - DEPT FIS	225 (5)	1.8 (9)	875 (6)	7 (14)	125
UAM I - DEPT QUIM	65 (16)	1.32 (30)	291 (14)	5.93 (19)	49
UAM I - DEPT MATEMAT	8 (41)	1.33 (29)	31 (48)	5.16 (26)	6
UAM I - DEPT IPH	8 (41)	1 (49)	36 (45)	4.5 (32)	8
BUAP					
BUAP - IF BUAP	167 (8)	1.65 (13)	547 (9)	5.41 (23)	101
BUAP - FCFM	68 (15)	1.23 (34)	183 (20)	3.32 (38)	55
BUAP - IC BUAP - CIDS	29 (27)	1.45 (20)	84 (36)	4.2 (33)	20
IPN					
IPN - ESFM	113 (10)	1.28 (32)	264 (17)	3 (44)	88
IPN - ESIQIE	26 (30)	1.18 (38)	50 (42)	2.27 (51)	22
IPN - ESIME	16 (35)	1.33 (29)	34 (46)	2.83 (46)	12
IPN - CICATA	9 (40)	0.9 (54)	18 (49)	1.8 (52)	10
UASLP					
UASLP - INST FIS	83 (14)	1.66 (12)	476 (10)	9.52 (5)	50
UASLP - IICO	42 (22)	1.61 (15)	154 (22)	5.92 (20)	26
UASLP - INST MET	9 (40)	0.9 (54)	10 (52)	1 (57)	10
UASLP - FAC CIENCIAS	8 (41)	1.14 (42)	43 (43)	6.14 (17)	7

* For a full description of the departmental acronyms see table A1 in appendix A.

Table 2.7. Total Number of Groups by Number and Type of Institutions

Number of Groups	All Institutions*	Top 25 Mexican Institutions
Multiple institutions**	297	291
Single institution***	231	191

* Includes the groups of all institutions, Mexican and non-Mexican, top and non-top
** Groups could belong to more than one institution
*** Groups with at least three researchers that belong to the same institution

Table 2.8. Summary Statistics for Groups at the Top 25 Mexican Institutions with a Single Institution

N=291	Total Number of (per RG)			Normalized by Researcher	
	Researchers	Articles	Citations	Articles	Citations
Max	26.0	81.0	380.0	10.1	62.0
Min	3.0	2.0	0.0	0.7	0.0
Average	4.8	12.9	56.5	2.6	11.1
STDEV	2.9	11.7	69.9	1.7	11.8

Table 2.9. Institutional Summary Statistics for Groups with a Single Institution

	Total (per Institution)				Average (per RG)				
	RGs*	Res**	Art	Cit	Res**	Art	Cit	Art/Res	Cit/Res
BUAP	14	70	166	596	5	11.9	42.6	2.2	8.4
CICESE	9	31	55	294	3.4	6.1	32.7	1.6	7.8
CIMAT	1	4	20	248	4	20	248	5	62
CINVESTAV	28	165	449	1748	5.9	16	62.4	2.7	11.6
CIO	9	46	128	807	5.1	14.2	89.7	2.7	15.1
INAOE	6	27	74	419	4.5	12.3	69.8	2.7	14.8
ININ	4	16	28	36	4	7	9	1.9	2.4
IPN	10	35	65	152	3.5	6.5	15.2	1.7	4.2
UAM A	2	6	18	63	3	9	31.5	3	10.5
UAM I	10	48	150	642	4.8	15	64.2	3.4	13.6
UNAM	74	356	991	4644	4.8	13.4	62.8	2.6	12
UANL	1	3	3	6	3	3	6	1	2
UAQ	3	10	15	55	3.3	5	18.3	1.4	5.1
UASLP	9	43	84	430	4.8	9.3	47.8	1.8	9.8
UAZ	1	6	18	58	6	18	58	3	9.7
UniGDL	2	6	24	54	3	12	27	4	9
UniGuan	3	15	94	348	5	31.3	116	5.2	18.5
UniMich	1	3	6	2	3	6	2	2	0.7
UniSon	4	18	67	181	4.5	16.8	45.3	3.9	10.7

* Research Groups from top 25 Mexican institutions with only one institution represented in each group.
** Res = Researchers

Figure 2.8 shows the distribution of the share of research groups (identified by the method) in an institution by percentile. The figure is built by ranking the groups based on the number of articles normalized by group size, and dividing this rank into percentiles. The assumption in this figure is that if the groups at each institution were drawn from a random process, each institution would have 10% of their groups on the 10th percentile, another 10% on the 20th particle, and so on. From this picture it can be seen that the system is quite heterogeneous within institutions, e.g. some institutions (like UNAM, CINVESTAV and UAM-I) have groups in almost all percentiles while other have only in a few brackets (like UniSon, INAOE and CICESE). In addition, there is some level of skewness towards a specific percentile. For example, BUAP does not have any RG at the top percentile and has two times more groups on the 50th bracket, whereas more that 40% of the groups at CICESE are at the lowest percentile.

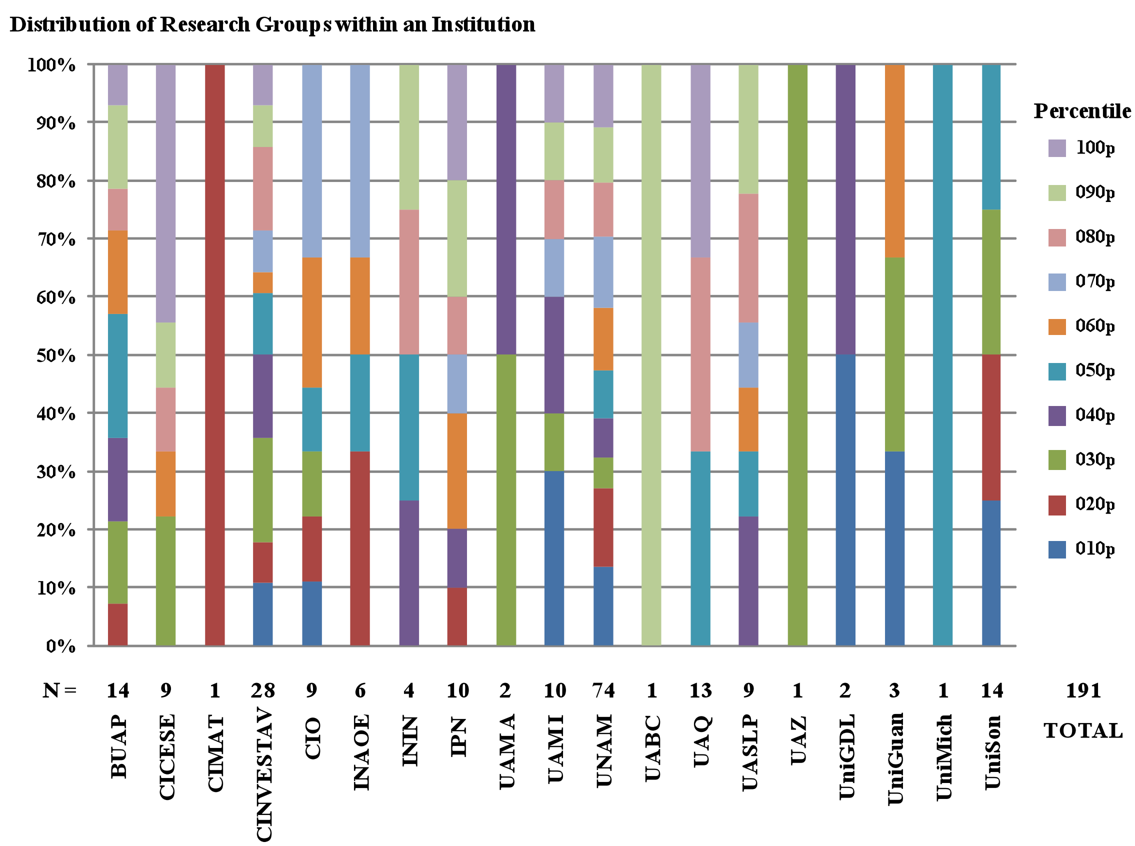
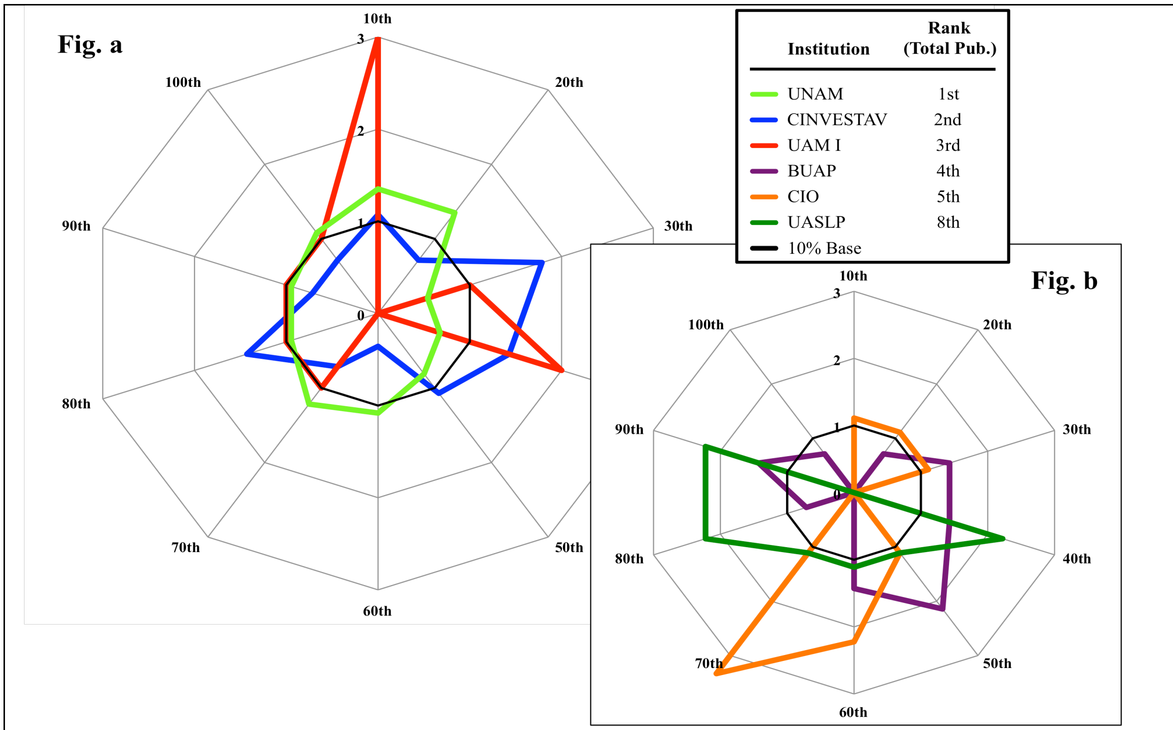


Figure 2.8. Distribution of Research Groups within an Institution by Percentile (based on publications per group adjusted by size).

Figures 2.9a and 2.9b show the heterogeneity within and across institutions by charting the relative strength (based on publications per group adjusted by size) each major institution has against the 10% base. The rationale is the same as in the previous graph. One would expect that all institutions will have 10% of their groups in the top 10 percentile, another 10% in the 20th bracket and so on; or in other words the shape of the decagon would be symmetric. For this picture it can be seen that the top three publishing institutions – UNAM, CINVESTAV and UAM-I (figure 2.9a) – have an asymmetric performance, e.g. UAM-I has 30% of its groups in the top percentile and none in the second cohort. In addition, UNAM performs better than CINVESTAV at the upper percentiles (10th and 20th) but trails this institution in the next three percentiles. Furthermore, this asymmetry in performance is particularly prevailing in BUAP, CIO and UASLP (figure 2.b).



Figures 2.9a. and 2.9b. Institutional Strength (based on publications per group adjusted by size) for Three Leading Institutions.

Figure 2.10 shows that this heterogeneity is also present in the system if the groups are ranked based on citations per group adjusted by size.

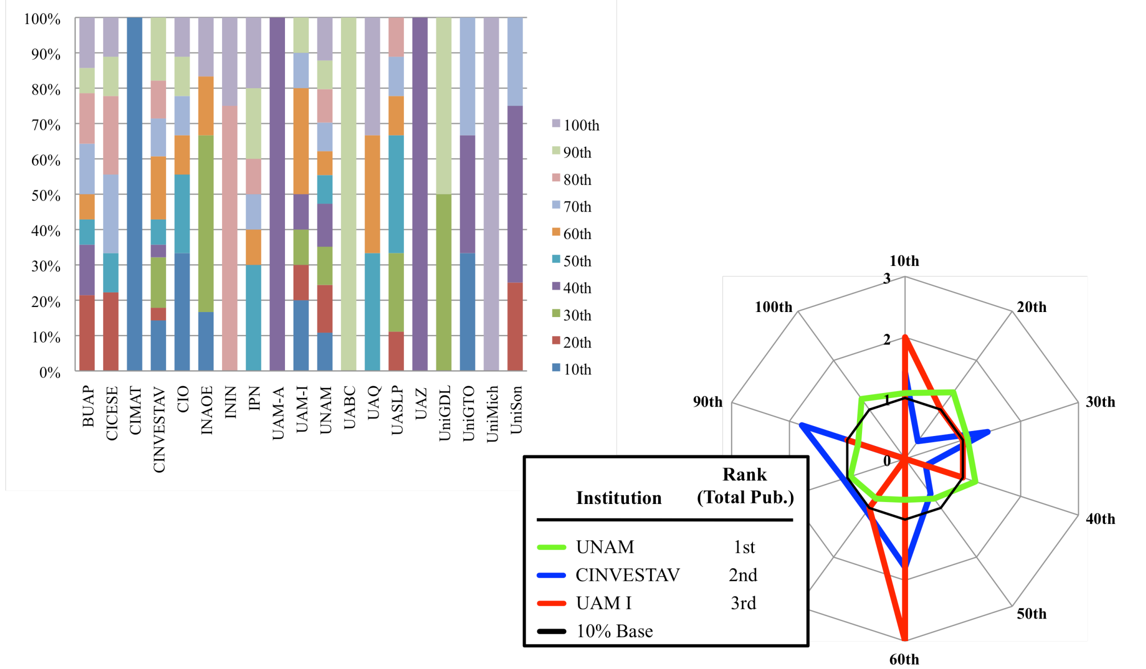


Figure 2.10. Institutional Strength (based on citations per group adjusted by size) for Three Leading Institutions.

As suggested in this analysis, identifying research groups should allow a much better understanding of the dynamics of scientific productivity within and across institutions and departments. Yet, this process also suggests that these groups will have very different characteristics with regard to the nature of their research. Therefore, it may not be reasonable to compare groups across the board, but rather complement the ideas described above with a method that would allow an identification of other benchmark groups with a comparable research profile. As described before, the proposed method allows us to compare a given group with others that produce comparable research by looking at the overlap of their KFP (or co-citations between the papers they publish) and varying their level of similarity between KFPs.

Tables 2.8 and 2.9 provide an example for research groups RG_037-UniSon and RG_049-CINVESTAV of how individual group benchmarks can be done. Overall, the first group has an average performance (in terms of publications adjusted by size) and the second has an average-to-low performance. But if we use the KFP to measure the distance between groups and focus only on their relevant peers, a somewhat different picture emerges. Group RG_037-UniSon improves its relative standing going from the 30th bracket to the 1st place at the 25% level of similarity (table 2.10). This happens because there are only four peer groups that share 25% backward citations and, within those, RG_037 has the top output performance. In contrast, group RG_049-CINVESTAV still has an average-to-low performance at the same level of similarity, ranking in 5th place, out of six peers (table 2.11).

Table 2.10. Relative Ranking Based on Group's Knowledge Footprint Over 5 Years, RG_037

		Level of Similarity	Articles per Researcher	Overall percentile	Rank within cohort			
					5%	10%	15%	25%
	RG_000-UNAM	10%	5.4	10th	1	1		
	RG_084-BUAP	21%	4.3	20th	2	2	1	
	RG_021-UniSon	10%	3.8	20th	3	3		
=>	**RG_037-UniSon***	**BASE**	**3.7**	**30th**	**4**	**4**	**2**	**1**
	RG_200-UAM-I	9%	2.3	40th	5			
	RG_000-CINVESTAV	10%	2.1	50th	6	5		
	RG_037-U. Auto Madrid**	100%	2		7	6	3	2
	RG_021-U. Sao Paulo**	10%	2		8	7		
	RG_000-Slovak Acad Sci**	10%	1.9		9	8		
	RG_000-IPN	10%	1.8	70th	10	9		
	RG_037-UNAM	100%	1.8	70th	11	10	4	3
	RG_084-U. Connecticut**	21%	1.3		12	11	5	
	RG_040-UNAM	6%	1	90th	13			
	RG_037-Hung. Acad Sci**	100%	0.7		14	12	6	4

*This group has 31 peers at the 1% level of similarity, 13 at the 5%, 11 at the 10%, five at 15% and three at 25%.
**These groups don't appear in the overall ranking for articles and citations, 4th and 7th column, because they are based in a foreign institution.

Table 2.11. Relative Ranking Based on Group's Knowledge Footprint Over 5 Years, RG_049

		Level of Similarity	Articles Per Researcher	Overall percentile	Rank within cohort			
					5%	10%	15%	25%
	RG_004-UniSon	34%	6.3	010p	1	1	1	1
	RG_000-UNAM	5%	5.4	010p	2			
	RG_007-CINVESTAV	17%	3.6	030p	3	2	2	
	RG_004-CINVESTAV	34%	3.1	030p	4	3	3	2
	RG_004-BUAP	34%	2.7	040p	5	4	4	3
	RG_004-UAQ	34%	2.3	050p	6	5	5	4
	RG_000-CINVESTAV	5%	2.1	050p	7			
	RG_078-CINVESTAV	15%	2	060p	8	6	6	
=>	**RG_049-CINVESTAV**	**100%**	**1.9**	**070p**	**9**	**7**	**7**	**5**
	RG_000-Slovak A. Sci**	5%	1.9		10			
	RG_000-IPN	5%	1.8	070p	11			
	RG_004-CICESE	34%	0.7	100p	12	8	8	6

*This group has 36 peers at the 1% level of similarity, 11 at the 5%, seven at the 10% and 15% and five at 25%.
**These groups don't appear in the overall ranking for articles and citations, 4th and 7th column, because they are based in a foreign institution.

Number of groups by level of similarity

In addition, the KFP can be used to assess how the number of relevant peers changes by varying the level of similarity between groups. Figure 2.11 gives the distribution of the number of peers by level of similarity; the x-axis has the different groups (identified with the method) and the y-axis has their respective number of peer groups at 1%, 5%, 10%, 25% and 50% level of similarity. From this chart it can be seen that the number of peers drops significantly from 9.5 peers on average (at 1% level of similarity) to 1.2 peers (at the 50% level of similarity). These results suggest that groups in these fields are in niche areas because the overlap of their knowledge footprint is low at high levels of similarity.

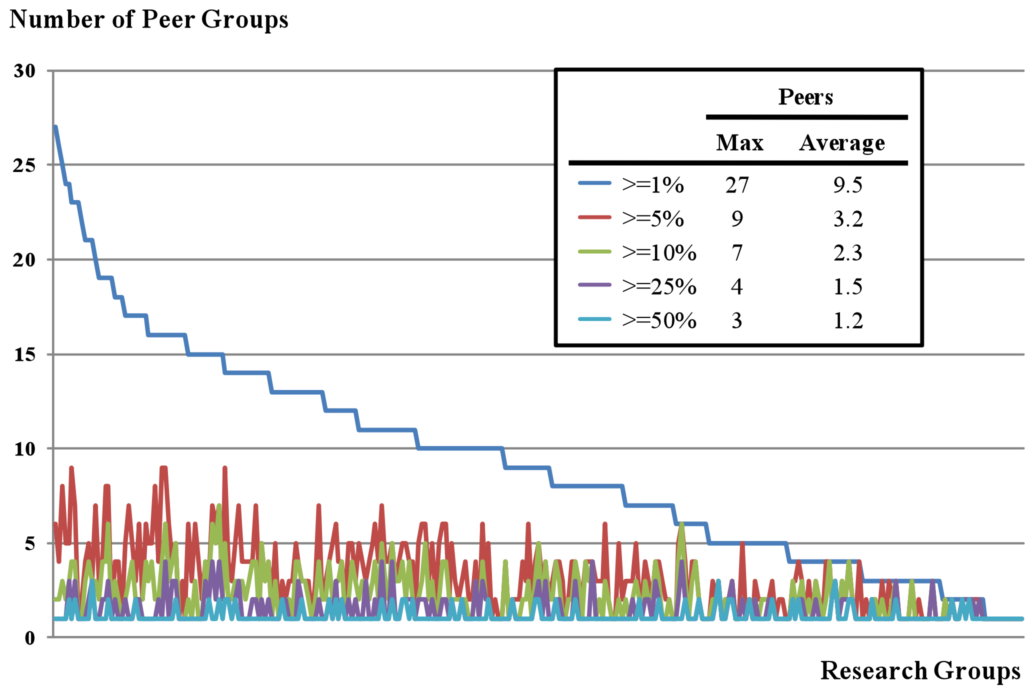


Figure 2.11. Distribution of Number of Peers by level of Similarity. This figure shows how the number of relevant peers changes by varying the level of similarity in the KFP. The x-axis has the different groups (identified with the method), the y-axis has their respective number of peer groups at different level of similarity, and each line represent the distribution of relevant groups. For example, the blue line (which is the baseline for ordering the groups in descending order) shows the distribution of these groups at 1% level of similarity, whereas the aqua line presents this distribution at the 50% level. In addition, this figure provides in the upper right hand the maximum and average number of peers at different levels of similarity. These results suggest that groups in physics (and related fields) in Mexico are in niche areas because the overlap of their knowledge footprint is low at high levels of similarity

The example above shows the power of combining a method that allows an identification of research group boundaries with another that characterizes distances in their knowledge footprint. The potential is for a very sharp and clear identification and benchmark of the relevant pockets of knowledge generation and impact in an institution or region, allowing a more precise evaluation and reward process for university administrators, program managers or policy makers.

2.5.3. Group Formation and Evolution

In addition to identifying and characterizing groups for a fixed period of time, the proposed method can also be used to understand how groups are formed, how they evolve, how their composition changes over time and the influence past performance has on current rankings. In this section we offer a brief description of how this method can be used to chart the evolution of groups over time.

First, we show how groups evolve and how their ranking changes over time. Table 2.12 shows the composition of four RGs in terms of authors (lines in the table) and how they evolve 1990-1994 (period 0) to 1995-1999 (period 5). At period 0 these groups stand at the top 10th percentile (with nine additional groups). Yet, over the course of five periods (each with a group id and their percentile) we find that some remained unchanged in terms of performance (RG 1 from Period I to Period 4), others split (RG 23 in period I), split and join (RG 36 in period I and RG 27 in period 3), or drop performance (RG 28 in period 2). In addition, we can assess the composition of teams in terms of group members and how this composition changes over time. For example, group 23 at period 0 (third group in the second column) has four members (A07 to A10); in period 1 this group breaks into two teams (RG9 and RG28) and each team adds one member (authors B01 and B02); in period 2, RG28 evolves into group 16 by adding authors A08 and B01, and researcher A07 creates a new team (RG12) with C01 and C02; by period 5, RG23 (from period 0) has evolved into two teams, RG39 and RG17, and there are only two remaining scientists from the original group, A08 and A10.

Table 2.12. Evolution of Research Groups

	Period 0 (90 – 94)	Period 1 (91 – 95)		Period 2 (92 – 96)		Period 3 (93 – 97)		Period 4 (94 – 98)		Period 5 (95 – 99)	
uthor	RG*	RG	Percentile	RG	Percentile	RG	Percentile	RG	Percentile	RG	Percentile
01	1	1	10	2	10	2	10	2	10		
02	1	1	10	2	10	2	10	2	10		
03	1	1	10	2	10	2	10	2	10		
04	21	23	10	33	10	50	10				
05	21	23	10	33	10	50	10	171	10		
06	21	23	10	33	10	50	10	171	10	561	NA
)1								171	10	561	NA
)1				12	20	16	10	23	10	39	10
)2				12	20	16	10	23	10		
07	23	9	20	12	20	16	10	23	10	39	10
08	23	9	20	16	40	18	40	29	40	39	10
)1		9	20	16	40						
)2		28	60	16	40	18	40				
09	23	28	60	16	40						
10	23	28	60	16	40	18	40	29	40	17	70
01						18	40	29	40	17	70
)1										17	70
)3		27	10	37	10	27	20	20	20	13	10
11	36	27	10	37	10	27	20	20	20	13	10
12	36	27	10	37	10	27	20	318	NA	525	NA
13	36	65	10	41	10	27	20	20	20	13	10
)4		65	10	41	10						
)5		65	10	41	10						
14	36	179	90	216	100	27	20	20	20		
)6		179	90	216	100	27	20	20	20		
)7		179	90	216	100	27	20	20	20		
)8		179	90	216	100	27	20				

ix out of 13 groups were identified at the top percentile in the period 1990-1994.
A = these groups were not taken into account because they only published one paper or had fewer than three researchers in the period.

In contrast to this, we can reverse this procedure to assess the way groups form over time and evaluate where the "pedigree" of a creating group came from. Table 2.13 shows how three groups (out of 31 at the 10th percentile) are formed between 1990-1994 (period -5) and 1995-1999 (period 0). For example, RG13 (the second group in the second last column of table 13) in period 0 was created with members from groups 12, 36, 58 and 125 from period -5. In addition, the data suggest that the current standing of RG13 (at period -0) emerged from the strength of groups 36 and 12 from period -5.

Table 2.13. Formation of Research Groups

Period -5 (90 – 94)		Period -4 (92 – 96)		Period -3 (94 – 98)		Period -2 (90 – 94)		Period -1 (92 – 96)		Period 0 94 - 98	
RG	Percentile	RG	Percentile	RG	Percentile	RG	Percentile	RG	Percentile	RG*	Authors
										9	A01
						12	10	17	10	9	A02
11	30	14	20	11	20	12	10	17	10	9	A03
								32	10	9	A04
333	NA	396	NA	435	NA	201	NA	32	10	9	A05
						305	NA	32	10	9	A06
								129	NA	9	A07
										13	A08
								20	20	13	A09
12	20	11	20	14	20	22	20	20	20	13	A10
						27	20	20	20	13	A11
36	10	27	10	37	10	27	20	20	20	13	A12
				41	10	27	20	20	20	13	A13
		65	10	41	10	27	20	20	20	13	A14
36	10	65	10	41	10	27	20	20	20	13	A15
125	30	65	10	41	10	27	20	20	20	13	A16
				81	40	27	20	20	20	13	A17
12	20	54	30	23	40	28	30	20	20	13	A18
58	50	51	40	81	40	41	50	20	20	13	A19
		309	20	502	20	521	10	20	20	13	A20
12	20	11	20	14	20	22	20	21	10	13	A21
				502	20	521	10	108	20	13	A22
								21	10	15	A23
		11	20	14	20	22	20	21	10	15	A24
12	20	11	20	14	20	22	20	21	10	15	A25
12	20	23	10	33	10	22	20	21	10	15	A26

* Three out of 31 groups were identified at the top percentile in the period 1995-1999.
NA = these groups were not taken into account because they only published one paper or had fewer than three researchers in the period.

2.6. Discussion and Policy Implications

In the last thirty years the realm of science and technology has evolved dramatically. These changes have fostered an evaluation culture (and industry), enhanced traditional methods (like peer review, Guston, 2003) and created new ones, including benchmarking between countries (May, 1997; Adams, 1998; King, 2004; Veloso et al., 2005) or socioeconomic assessments (van Raan, 2000). While useful, these evaluation methods have an important common limitation: the boundaries of the focal unit are typically artificial and rigid, failing to notice unique and self organizing characteristics of the research endeavor.

To address this limitation, this paper proposes an evaluation method that takes into account the endogenous, or self-organizing, characteristics of research groups. It defines research groups based on the strength and frequency of the collaboration patterns (within a field of knowledge) and ranks them using the level of similarity of their knowledge footprint (i.e. common citations). In addition, this method is tested with a database from the fields of Physics and Applied Physics/Condensed Matter/Materials Science and Optics in Mexico containing all the papers published between 1995 and 1999 (as reported by ISI). A detailed full and relative peer benchmark is performed for both areas.

The method developed in this paper and applied to the context of Mexico produces three main results. First, as expected, the strength and frequency of the collaboration patterns allows us to single out cohesive groups, i.e. this method identifies the key research groups (or collective actors) in a field of knowledge, regardless of the institutional or location context of the members (researchers). In addition, this new technique allows scholars and policy makers to take into account the (expected) heterogeneity within institutions in their assessments. This is a departure from traditional methods because a potential evaluator would not normally be able to assess the internal cohesiveness of groups, or self-organizing mechanisms.

Second, the knowledge footprint (KFP) and the benchmark at different levels of similarity in KFP allows a departure from the established evaluation literature. This step allows potential evaluators to identify similar research groups, assess these groups and produce more meaningful comparisons and rankings (e.g. see tables 8 and 9). This solution contrasts with the more traditional approach, where the evaluator typically uses broad and artificial similarities, such as comparing mechanical engineering departments across universities, assuming that they are more or less the same. In addition, this method has an important feature: it can easily be extended to other types of focal units, including, institutions, departments, networks or regions, and even individuals, with minor modifications.

Third, the research done by the different groups in the areas of Physics, Applied Physics/Condensed Matter/Materials Science and Optics in Mexico is (almost) non-redundant. In fact, the KFP overlap of these groups is relatively small, which means that each RG is (virtually) focused on one area of the research space. For the casual observer, this could be seen as something good if resources are used efficiently to support only one RG in each area and there is no duplicity in teams. However, this could create a vicious cycle by creating a system where there is no competition and there are no emergent ideas. In order to overcome this, it would be healthy if international visiting committees assessed the performance of the RG and their performance compared with groups in similar economies, like Brazil.

From the preliminary results one can conclude that this method can support policy makers and scientist to better identify the frontier of research groups and to find suitable and relevant peers for benchmarking. In addition, this procedure allows scholars, as well as policy makers, to better understand the self-organizing mechanisms of research groups and assess how they evolve over time. We believe this whole process will increase our knowledge of the research endeavor and, combined with other methods (like peer review), will produce better assessments. In addition, this method helps to close the gap between performance analysis and the mapping of science in bibliometric analysis by giving a different perspective to Noyons, Moed, and Luwel (1999). It also creates a link between the areas of research evaluation and network analysis.

The development of this new method also generates new questions that need to be addressed in subsequent work. One of them is the effect of weakening the clique assumption, using other measures of group cohesiveness (e.g. n-cliques, k-plexes, etc.) to define collaborative groups. In addition, further analysis is also needed to test the robustness of this method, by incorporating other fields of knowledge or using data from other countries or regions.

Finally, this method could further our understanding of the determinants of research group productivity (Gonzalez-Brambila and Veloso, 2007) in a number of ways. One possibility is to study how the characteristics of the naturally emerging groups are tied to their productivity. Another possibility would be to extend the approach to other types of research output data amenable to equivalent analysis, in particular patents.

Chapter 3. Birth of prominent scientists

3.1. Introduction

Today's emphasis on economic activity based on knowledge and innovation is leading industrialized as well as developing nations to place an important emphasis on policies to advance their science, technology and innovation (ST&I) systems and reap their benefits (OECD, 1999, 2001, 2004a, 2010). At the core of these efforts are policies to expand the scientific base and to generate, attract and retain highly talented scholars (Cervantes, 2004). Nations around the world have been pursuing a variety of strategies to this effect (Laudel, 2005). The most common approach is an effort to grow the size of their research system, aiming to build a critical mass of researchers across a variety of areas. For example, the Mexican Government has enlarged its scientific system by funding the training of scientists (with national and international fellowships) and developing repatriation and post-doctoral programs for researchers from Mexico27. Some nations have placed an emphasis on attracting and retaining some of the world's most accomplished and promising minds (Urquhart, 2000; Australian Research Council, 2001; Pickrell, 2001). A good example of such strategy is Canada's Research Chairs Program28. The underlying assumption is that these key scientists can play a vital role in the development of a research system because they will make groundbreaking scientific discoveries, as well as create and develop internationally renowned research centers, improve universities' capacity for generating and applying new knowledge, train the next generation of highly qualified personnel and also enable the establishment of successful high-technology startups. Consequently, there is

[27] See programs for graduate studies and support for scientific research in http://www.conacyt.gob.mx/
[28] http://www.chairs-chaires.gc.ca/

growing interest within research administrators, policy makers and scholars in the role scientific stars have on the development of an ST&I system.

This has motivated an important stream of research focused on quantifying the impact leading researchers have on an established system, their peers and the institutions they work for. According to Zuker and Darby (1998) 0.8% of the scientists in the GenBank in the 1990s were 22 times more productive than the average scientist, publishing 17.3% more papers. In addition, Azoulay, Zivin and Wang (2007) as well as Oettl (2009) have shown the impact that “superstars” have on their peers by calculating the drop in productivity when a leading scientist dies. In the first study, Azoulay and his colleagues show that coauthors of an ‘extinguished’ star “suffer a lasting 8 to 18% decline in their quality-adjusted publication output;” whereas Oettl documents a higher loss in productivity, between 19 to 35%. In a related study, Goodall (2009) shows that accomplished scholars appointed as presidents (vice chancellors) of a university have a positive impact on the research quality of their institutions. Furthermore, Zucker and Darby (2010) find that stars themselves, rather than the disembodied knowledge associated to them, are crucial for the entry of a broad range of high-tech startups.

While stars may be important, collaboration networks, or teams, is another dimension considered to be of paramount importance in the development of a research system. A variety of authors have looked at how teams condition scientific production. One well-established dimension is that research collaboration has been on the rise in the last decades29 (Beaver, 2001; Wagner and Leydesdorff, 2005). This suggests that the teamwork research model rather than an individual-based approach is now the norm in most scientific endeavors (Adams et al., 2005;

[29] This increase has been uneven even across fields (Newman, 2004; Bukvova, 2010).

Wuchty, Jones and Uzzi, 2007). The idea is that a collaborative approach in the production of knowledge enables scientists to access complementary expertise (Katz and Martin, 1997; Melin, 2000; Beaver, 2001), valuable equipment and resources (Melin, 2000; Beaver, 2001), while exposing them to new ideas and encouraging cross-fertilization across fields (Beaver and Rosen, 1978, 1979a,b; Katz and Martin, 1997; Melin, 2000). This change also appears to have had a positive effect on publishing productivity (Melin, 2000; Lee and Bozeman, 2005), quality (Persson et al., 2004; Rigby and Edler, 2005; Wuchty et al., 2007; He et al., 2009), as well as visibility and prestige (Crane, 1972; Beaver and Rosen, 1978, 1979a,b; Katz and Martin, 1997; Beaver, 2001).

Some authors have also considered the extent to which ensembles of scientists provide a nurturing environment where researchers can flourish, in particular younger ones. For example, Bozeman and Corley (2004) emphasize the importance of collaboration on the development of the scientific and technical human capital30 of researchers, especially when a senior scientist works with a junior one and the former acts as her mentor. According to these authors, under the right circumstances, a graduate student or post-doctoral researcher can gain, "not only enhanced S&T knowledge, but craft skills, know-how, the ability to structure and plan research and, of course, increase contacts with other scientists, industry, and funding agents." Oettl (2009) complements this perspective, noting that helpful eminent scientists have a greater impact (between 58% to 84% more) on the productivity of their co-authors than just highly productive researchers.

[30] Bozeman et al (2001) defines "Scientific and technical human capital" (S&T human capital) as "the sum of researchers' professional network ties and their technical skills and resources."

Research focused on the effects of sponsorships in academia has assessed the extent to which science follows Merton's norms (1974)31 by trying to disentangle the impact achievement and ascription (like race, gender, institutional affiliation or recognition of close collaborators) have on scientific performance, allocation of resources and academic career success. This exploration has generated some evidence that location (e.g. institutional affiliation or doctoral origin) influences the reward structure of science. For example, Crane (1965) showed that the caliber of the institution has a positive impact on the productivity and prestige of its scientists, as well as the rate of the number of grants (Cameron, 1981; Long, 1978). Allison and Long (1990) suggest that changes in productivity can be ascribed to changes in departments and the prestige of these entities, while Reskin (1977), as well as Long and McGinnis (1981), establish that organizational context (like industry or academia) can influence a scientist's level of performance. In addition, work focused on individuals has shown that the sponsor's talent also matters. For example, Long's (1978) study of academic biochemists found that sponsors' citations affected their students' number of publications and citations, as well as their prestige; and Reskin's (1979) work on doctoral chemists showed that the advisor's productivity influenced the pre-doctoral productivity of its advisees. Furthermore, advisees have a propensity to follow the steps of their advisors by replicating their success and skills. For example, Zuckerman (1967) showed that Nobel Prize winners tend to positively influence the chances of their students and collaborators in also becoming Nobel Laureates; and, in a recent study, Malmgren et al. (2010) found that protégés that were trained by high fecundity32 mentors also score high on this

---

[31] Merton (1973) argues that science should be governed by the norms of universalism, communism, disinterestedness, and organized skepticism; implying that a successful academic careers should be based on talent and not be determined by ascriptive characteristics.
[32] Malmgren et al. (2010) define *mentorship fecundity* as the number of protégés a mentor trains.

indicator. Finally, there is a set of studies that have assessed the extent to which science is stratified by race and gender (e.g. see Long and Fox, 1995; Levin and Stephan, 1998; Fox, 2001).

Despite the advances in our understanding of how the research context impacts the development of scientists, old and new, much remains to be explored. First, work focused on quantifying the influence of mentors and eminent scientists on others has not considered how they interact with other scholars in the context of research teams, and their impact on the evolution of the system, especially in terms of others at the beginning of their career. Second, previous research is typically composed of case studies that recount the mentoring experience, cross sectional studies or longitudinal analysis with usually short time frames as well as small and often random data sets (e.g. Long et al., 1979; Reskin, 1979; Green and Bauer, 1995; Williamson and Cable, 2003; Judge et al., 2004; Paglis et al., 2006). Third, with the exception of Malmgren et al. (2010), research on mentorship has not looked at the extent to which protégés mimic their mentors' steps, performance and reputation.

Furthermore, because the Science and Technology (S&T) community has different characteristics around the world (Nelson, 1993), a better understanding of the factors that condition research output, impact and success in science requires an analysis of a diverse set of countries. S&T systems have particular disparities between developing and developed nations. In developing nations there are fewer resources and less infrastructure dedicated to research and development (R&D). Moreover, the government funds most R&D and human, as well as financial, resources are centralized in a few institutions33. Thus, studying emerging economies

[33] E.g. in Mexico in 2002 68% of the Gross Domestic Expenditure on Research and Development (GERD) was financed by the public sector (CONACYT, 2004, p16). In addition, in 2003 the National Autonomous University of Mexico (UNAM) had 27% of all the researchers belonging to the National Research System (SNI,

provides a better understanding of the factors that influence the performance, impact and overall contribution of scientists in this environment (Nelson, 1993). In addition, studying this type of countries is relevant because these are actively developing and implementing policies to improve their S&T systems. Therefore, a better understanding of the factors that foster success at individual and aggregated levels could help leap forward their system. This is particularly pertinent because, with a few exceptions (Veloso, et al. (2006); Gonzalez-Brambila and Veloso (2007). Wagner (2008); Ordonez-Matamoros et al. (2009); Horta, et al. (2010), chapters 2 and 3), research in this area has mostly focused on the developed world.

This research tries to bridge existing gaps by combining the different research streams described above, with two complementary dimensions. One is to look at the role that scientific stars (i.e. the most accomplished and salient researchers) have in a science system. In particular we assess how relevant these eminent scientists are for the development of a system. This means understanding how much they contribute to the output and impact of the system, as well as how influential they are in breeding the next generation of successful scientists, i.e. how successfully their protégés mimic their stellar performance. The other dimension is to assess how collaboration conditions the development of incoming scientists. In particular, we will look at the importance of the collaboration network of early co-authors for the productivity of new scientists and the likelihood that they also become leading scientists. Furthermore, this study uses a unique data set that spans almost two decades and allows us to look at the research system of a developing country, Mexico.

This work is divided in four sections. In the first section we lay out the purpose of this

---

http://www.conacyt.gob.mx/sni/) and received almost 50% of the federal R&D funding, and four public institutions monopolized 92% of this budget (CONACYT, 2004, p24).

study and its research questions. In the second we explain the methods we are going to use, explain the data and define some key concepts. The third part provides results. Finally, we provide conclusions and policy implications.

3.2. Research Questions

Previous research on the impact that (lead) scientists and mentors/sponsors have on other researchers has focused on quantifying the direct influence the former have on the productivity of their collaborators (Azoulay, Zivin and Wang, 2007; and Oettl, 2009), or in assessing how they contribute to successful postdoctoral careers, including their first job and subsequent awards (Reskin, 1979; Green and Bauer, 1995; Judge et al., 2004; Paglis et al., 2006). In this study we also consider the importance of accomplished scholars in a scientific system. Yet, instead of quantifying the impact these scientists have on ongoing relationships we measure the influence they exert on young faculty when they first enter a particular field. In addition, we measure the impact different nurturing environments (or research collaborations) have on the incoming scientist. Furthermore, we look at the extent to which new scientists follow the steps of their mentors and also become a star.

3.2.1. Overall contribution of eminent scientists to a science system

The notion that a small percentage of researchers contribute to a disproportionate share of output in terms of papers (Lotka, 1926; Pirce, 1963; Zucker, Darby, and Brewer, 1998) and citations (Hagstrom, 1968; Cole, 1970; Cole and Cole, 1972; Allison and Stewart, 1974) is well established in the literature. For example, Pirce (1963) found that a minority of scientists in physics (around six percent) publishes 50 percent of all the publications, while Cole (1979) and Reskin (1977,1978) have shown that this percentage of contributing scientists is fifteen percent in several other fields[34]. Allison and Stewart (1974) also found that the distribution of citations is more unequal than the one for articles and that this inequality increases with tenure, for both measures. In addition, highly accomplished researchers also influence the realm of science by training, coaching and working with the next generation of eminent scientists, contributing indirectly to the system through the work of their advisees. For instance, Zuckerman (1967) found that 62% of the Noble laureates (in his sample) worked as young researchers under the supervision of previous prize-winners; and these eminent scientists were more inclined to collaborate with other distinguished and highly productive researchers than their less renowned counterparts. Furthermore, this skewness in productivity and impact can be more pervasive in emerging economies where limited resources and heterogeneity within the system (as seen in chapter 2 of this thesis) might favor a few scientists. With this in mind in this study we consider the direct impact that eminent scientists have in the system, by quantifying their total output and citations, as well as the number of subsequent stars that they breed. In addition, we also assess the indirect contribution these researchers have in the system by looking at the performance of the scientists they breed.

3.2.2. Mentorship, Research Environments, Apprenticeship and Performance

As previously stated, mentors can play an important role in the development of their protégés, having a positive impact in their careers (Phillips, 1979; Reskin, 1979; Bargar and Mayo-Chamberlain, 1983; Kram, 1985, p. 8; Cronan-Hillix et al., 1986; Fagenson, 1989; Green, 1991; Cable and Murray, 1999; Tenenbaum et al., 2001; Allen et al., 2004; Paglis et al., 2006). Within academia, doctoral students are more than protégés. They are typically apprentices of

---

[34] Cole (1979) showed this on natural, biological, and social sciences, whereas Reskin (1977, 1978) proved it for chemistry.

their advisors and sometimes part of the research group of their mentors. This group provides direct input into the work performed by a student, showing her how to conduct research and get published (Reskin, 1979; Judge, 2004). Students working under the supervision of an eminent scientist could have additional benefits by being exposed early on in their careers to promising research ideas and being able to interact and collaborate with other reknown researchers, including Nobel Laureates (Zuckerman, 1967). Ham and Weinberg (2007) analysis on Nobel laureates showed that being surrounded by other prize-winners had a significant positive effect on starting their own work that would yield this type of recognition. With this in mind, we will study how much the productivity of a new researcher increases if he or she enters into the system by the hand of a star.

While the presence of stars is likely to be important, it is clearly not the only important aspect in the research environment of a nascent scientist. The organizational context and research/collaboration environment where scientists do their work are also likely to play an important role in fostering or hindering the productivity of budding researchers, in particular graduate students and postdoctoral fellows (Fox, 1983; Fox and Mohapatra 2007; Louis et al. 2007). For example, Long and McGinnis (1981) found that an appointment in non-intensive research organizations depresses publication output, whereas employment in research universities fosters publication. Work at the department level has also found that scientists publish more when they are surrounded by productive peers (Braxton, 1983) and research-oriented coworkers (Bland and Ruffin, 1992). Furthermore, leadership within research organizations of accomplished and experienced scientists is an additional factor that can affect productivity. As Dill (1992) noted, a team leader is there to “to influence member’s knowledge and values, to facilitate contact and networks, to attract other competent researchers, to help colleagues who are blocked or stopped in their researcher efforts, and so on.” (Bland and Ruffin (1992) citing Dill (1990, 1992)).

Prior work highlighting the role of both academic stars and the collaborative research environment of a new scientist establishes the critical questions we are interested in exploring. In particular, we assess the extent to which different research/collaborative environments influence the performance of incoming researcher by considering four contexts in which a new researcher becomes active in a scientific system. First, we consider that a new researcher enters the system in the context of an established research group (RG). Second, we separate the top RG from the rest, recognizing that leading groups may have some different characteristics from the average research group. We then consider whether the early collaboration of the new researcher is with the leader of an RG vs. the mentoring of other members of the group. Finally, we consider early mentoring by the leader of top RG.

3.2.3. Following the steps of giants: Mimicking Success

Previous studies on eminent scientists have stated the importance of sponsorship of leading researchers in their success. For example, Zuckerman (1967) found that young scientists working with Nobel Laureates tend to replicate the success of their senior collaborators; and Crane (1965) showed that the best students work under the supervision of top researchers at leading schools and become the next generation's most productive scientists. In addition, the mentorship literature has noted that “the majority of participating mentors had been involved in a previous mentoring relationship as a protégé” (Allen el al., 1997), suggesting that some advisees follow the steps of their advisors. Yet, with the exception of Malmgren (2010), little has been done either to predict in advance those who will follow the same career path or to measure the likelihood of their becoming a successful researcher. This research will look at the likelihood that a new researcher becomes a star if she enters with one.

3.3. Method

3.3.1. Database

To answer the previous questions we will use a database from Thomson Scientific[35] (Institute of Scientific Information, 2003) containing all papers published between 1980 and 2003 with at least one address in Mexico. This database contains the following information: article name, author(s), author(s) address(es), year of publication, journal, volume, pages, backward citations (i.e. references) and total number of citations received.

From this database we selected all the papers published in Mexico in the areas of Physics and Applied Physics/Condensed Matter/Materials Science in the period of 1981-2003[36]. We chose these areas because in the past the files of Physics and its related areas have been widely studied around the world (e.g. Collazo-Reyes et al. ,2004; Shrum et al., 2007) and Mexico has a long tradition of publishing in international peer-reviewed journals, indexed by ISI in these areas (ISI, 2003; CONACYT, 2008).

Once all the papers were identified, we created a dataset containing the name of the article, its author or authors, the institutional affiliation of these scientists and the number of citations these articles received within a three-year window (e.g. for the papers published in 1990 we restricted the citation count to the period 1990-1992). In addition we divided this set into three periods. Period one (which includes all the papers published between 1981 and 1983) is

---

[35] Formerly known as the Institute for Scientific Information (ISI)
[36] We only considered articles; this means that letters, notes and reviews were excluded. In addition, the extracted data have undergone a detailed cleaning and then processed to bibliometric indicators.

used to identify the scientists that were in the system before the focal period used in the study. Period two (from 1984 to 2001) is the sample period and includes only scientists that entered the system, i.e. published an article for the first time, after 1984 up to 2000. Period three (2002-2003) was used to identify the scientists that exited the system before 2001 and entered after 1983. We say a researcher entered the system (within the focus period) if her name was not present in the first period but appeared (or published an article) after 1983; in the same respect, a researcher exited the system if she was present in period two but not in the third one. In order to avoid sporadic authors, we excluded from this dataset all the researchers that published only one paper and were present only one year within the sample period 1984-1999.

3.3.2 Definitions

For this analysis, we classify all the researchers in our sample along several dimensions.

Star and non-star scientists

First, we characterize the researchers on our sample as star (or eminent) and non-star scientists based on their performance for a certain period of time along two dimensions: productivity (measured in terms of papers per year) and impact (citations per year). Previous studies have used a 5% cutoff point in output or impact to define an elite group of scientists (Azoulay and Wang, 2010; Oettl, 2009). Although this characterization seems easy enough, it is difficult to decide whether to draw a precise cutoff point at a certain number like 1%, 5% or 10% because of the skewness of these variables. With this in mind, we follow a different approach and define this select group of researchers using the sample's performance distribution. In this study a star scientist is a researcher who is above the average productivity plus one standard deviation (STDEV) of all scientists in the sample and a non-star scientist is one who is below this threshold. We use one STDEV as the minimum performance level; however, more stringent levels (two or three STDEVs) can be used to identify these key people[37].

Research environments

Second, using a novel method for the characterization of research groups (see chapter 2 for a full description) we identify all the groups that are present in the system, as well as leaders of these communities; and based on this, single out different research environments to which researchers might be exposed.

The main idea behind this new method is that it defines all the research groups based on self-organizing characteristics of the research endeavor (Guimerà, 2005). The proposed method uses the notion that modern science is conducted primarily through a network of collaborators (or groups) who organize themselves around key researchers, often known as the principal investigators (PIs). Specifically, this method uses the patterns of collaboration and the strength of ties in a co-authorship network to, first, identify the PIs[38], or leaders, of these ensembles and, then, to characterize the boundaries of different research groups (RGs) centered on these key people (see chapter 2). This allows us to identify disjoint groups within a certain period of time, from two years to the entire period of study, by including all the papers that were published within those years. For this research we used a three-year rolling window. This meant that, first we characterized all the groups between t_0 and t_2, then we did it for t_1-t_3, up to t_{n-2}- t_n, where n is the maximum number of years of our sample. This method provides a list of all the PIs in the

---

[37] Below we show the number of stars by different STDEV levels and their average contribution and performance. In addition, we show a complete regression analysis of the <u>one</u> STDEV definition and the most important results for the two STDEV.
[38] This method defines a *Principal Investigator* (PI) as an author with a high number of repeated connections, i.e. a researcher that has written several papers with a high number of coauthors.

system and the composition of these communities.

Based on the output of this method (i.e. the PIs and composition of RGs) and the performance of these groups[39], we distinguish four types of research environments to which scientists might be exposed (at the beginning of their careers). Table 3.1 defines these environments.

Table 3.1. Definition of Research Environments

	Research environment	**Definition**
1.	Exposure to a research group	A scientist is said to be exposed to an RG if he or she collaborates with a researcher that belongs to a RG
2.	Exposure to a <u>top</u> research group	A scientist is said to be exposed to a top-RG if he or she collaborates with a researcher that belongs to a top-RG
3.	Exposure to the leader of a research group	A scientist is said to be exposed to the leader of an RG if he or she collaborates with the PI of an RG
4.	Exposure to the leader of a <u>top</u> research group	A scientist is said to be exposed to the leader of a top-RG if he or she collaborates with the PI of a top-RG

Tables 3.2 to 3.4 provide the summary statistics for the total number of papers, citations, authors and different type of stars for the 1981-2001 period. From table 3.2 it can be seen that 4,180 unique authors were identified and 2,018 (or 48% of the total) were defined as suitable ones; in order to avoid sporadic authors we restricted our analysis to scientists that published two or more papers between 1984 and 2001 in at least two different years and entered the system before 1999. In addition, we can observe that the latter published 6550 articles in the sample period. Table 3.3 shows how the 2,018 authors break into two groups: the ones that entered the system before 1984 (8%) and the ones that did it after 1983 (92%). In addition, this table shows the number of eminent scientists based on articles and citations for these periods of time for the one-STDev definition (11% and 9% respectively) and two-STDev definition (~4% for both

[39] Following a procedure similar to the one we used to define star and non-star scientists, we classify these communities into top and non-top research groups, a *top research group* is a group that is above the average productivity plus one standard deviation (STDEV) of all the groups in the sample, and a *non-top research group* is below this threshold.

categories). Furthermore, table 3.4 illustrates how much star scientists contribute (directly) to the system. Depending on how you define an eminent scientist (articles or citations) they publish 28% to 42% of all articles and receive roughly 50% of all citations for the one STDev definition.

Table 3.2. Summary Statistics, Absolute Numbers of Papers and Authors

	Authors		Papers	
Total number, 1981-2001	4180		7223	
Excluded form the analysis	2162[1]		673[2]	
Remaining number, 1984-2001	2018	48%[3]	6550	91%[3]

[1]Authors were excluded because they published only one paper, were present for only one year, exited the system before 1984, or entered the system after 1998.
[2]Papers were not included because they were published before 1984 or by one of 2,162 excluded authors.
[3]Percentage of the total number.

Table 3.3. Summary Statistics, Absolute Numbers of Stars

		Stars Based on							
		Articles per year (by star-articles)				Citations per year (by star-citations)			
		One STDev[1]		Two STDev[2]		One STDev[1]		Two STDev[2]	
Authors that entered before the studied period (1981-1983)	169	16	9%	9	5%	17	10%	7	4%
Authors that entered within the studied period (1984-1998)	1857	201	11%	81	4%	167	9%	71	4%
Remaining number, 1984-2001	2018	217	11%	90	4%	184	9%	78	4%

[1] One STDev = Stars are above the average plus one Standard Deviation.
[2] Two STDev = Stars are above the average plus two Standard Deviation.

Table 3.4. Direct Contribution[1] by Different Type of Star (One STDev)[2]

Papers			Citations		
by star-articles[3]	2740	42%[6]	by star-articles[3]	6368	50%[6]
by non-star-articles	5331[5]		by non-star-articles	9866[5]	
by star-citations[4]	1822	28%[6]	by star-citations[4]	6263	49%[6]
by non-star-citations	5766[5]		by non-star-citations	9435[5]	
Total number of papers, 1984-2001	6549		Total number of citations, 1984-2003	12770	

[1] Direct Contribution is the total number of published papers by a star, as well as the received citations of those papers.
[2] Stars are above the average plus one Standard Deviation.
[4] star-citations, stars are defined based on citations per year.
[3] star-articles, stars are defined based on articles per year.
[5] Non-star papers or citations could be double counted with those of stars, since a given paper might be published by both stars and non-stars.
[5] Percentage of the total number.

Table 3.5 gives a general overview of the evolution of the system in terms of number of researchers, stars and research groups (identified by the algorithm developed in chapter 2); total and average productivity (articles and citations), and average performance by type of stars and research groups for the three-year periods of 1984-1996 to 1999-2001 (the studied period) and 1981-1984 (as reference). From this table it can be seen that between 1984-1986 and 1999-2001 the system expanded six-times in terms of number of researchers and publications, as well as 4.6 times in terms of citations; growing 13.8%, 14.1% and 12.2% each year, respectively. In addition, the number of research groups grew 5.3 times and the average size increased by 50%. Furthermore, the number of stars grew at faster rates. During this period of time, individual performance steadily increased by any measure, but group efficiency declined on all variables.

Table 3.5. System Evolution, 1981-2001

	81-83[1]	84-86	87-89	90-92	93-95	96-98	99-01	Growth[2]		CAGR[3]	
								81-83 / 99-01	84-86 / 99-01	81-83 / 99-01	84-86 / 99-01
Number of researchers											
Total number	169	216	334	531	930	1,422	1,503	7.9x	6x	12.9%	13.8%
Stars, articles[4]	16	19	36	55	110	178	202	11.6x	9.6x	15.1%	17.1%
Stars, citations[5]	17	18	32	43	91	129	142	7.4x	6.9x	12.5%	14.8%
Number of articles											
Total number	245	303	408	658	1,164	1,820	2,197	8x	6.3x	12.2%	14.1%
by star, articles[4]	71	83	129	168	408	825	1,127	14.9x	12.6x	15.7%	19.0%
by star, citations[5]	67	76	116	128	296	521	685	9.2x	8x	13.0%	15.8%
Average number of articles per author											
Average total	1.9	2.0	1.8	2.0	2.1	2.5	3.2	0.7x	0.6x	2.7%	3.3%
by star, articles[4]	4.6	4.6	3.9	3.6	4.5	6.2	7.7	0.7x	0.7x	2.8%	3.5%
by star, citations[5]	4.2	4.9	3.9	3.5	4.1	5.5	6.7	0.6x	0.4x	2.5%	2.1%
Number of citations											
Total number	633	736	867	1,261	2,242	3,536	4,135	5.5x	4.6x	10.4%	12.2%
by star, articles[4]	212	246	308	424	1,047	1,985	2,358	10.1x	8.6x	13.5%	16.3%
by star, citations[5]	209	292	380	573	1,069	1,771	2,178	9.4x	6.5x	13.1%	14.3%
Average number of citations per author											
Average total	6.3	6.9	5.3	5.8	6.4	7.0	8.2	0.3x	0.2x	1.4%	1.1%
by star, articles[4]	16.1	15.3	10.0	11.2	14.4	17.2	19.3	0.2x	0.3x	1.0%	1.5%
by star, citations[5]	16.7	22.5	13.3	18.2	18.0	23.6	27.8	0.7x	0.2x	2.7%	1.4%
Research Groups (RGs)											
Total number of RG	32	36	52	101	173	271	226	6.1x	5.3x	11.5%	13.0%
Top RGs, articles[4]	8	5	7	14	24	44	34	3.3x	5.8x	8.4%	13.6%
Top RGs, citations[5]	3	7	5	12	18	37	28	8.3x	3x	13.2%	9.7%
Researchers per researcher group											
Average number	4.1	3.9	4.4	4.2	4.8	5.5	6.2	0.5x	0.6x	2.3%	3.1%
Average number of articles per researcher per group											
Total average	9.1	10.1	9.5	7.4	6.5	4.5	5.0	-0.5x	-0.5x	-3.3%	-4.6%
by top RGs, articles[4]	17.1	20.6	16.2	17.6	14.2	12.5	12.9	-0.2x	-0.4x	-1.6%	-3.1%
by top RGs, citations[5]	18.0	20.5	15.0	16.5	13.9	10.1	11.4	-0.4x	-0.4x	-2.5%	-3.8%
Average number of citations per researcher per group											
Total average	19.1	21.4	18.0	15.0	11.7	7.9	8.7	-0.5x	-0.6x	-4.3%	-5.8%
by top RGs, articles[4]	39.9	43.0	29.9	39.2	26.0	22.7	28.1	-0.3x	-0.3x	-1.9%	-2.8%
by top RGs, citations[5]	44.7	50.8	36.2	45.8	35.4	32.3	35.4	-0.2x	-0.3x	-1.3%	-2.4%

[1] Warm-up period.
[2] "x" denotes times of expansion or contraction.
[3] CAGR: Compound Annual Growth Rate.
[4] Based on articles, one STDev definition.
[5] Based on citations, one STDev definition.

Entry into the system

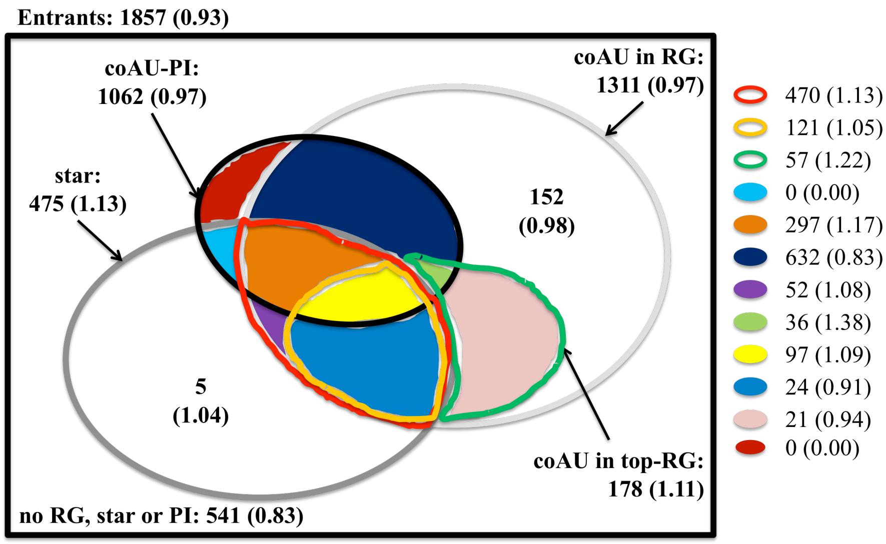
In addition to establishing if a researcher is a star (within a certain period) and the different type of environments this person has been exposed to during the early part of her academic career, we also identify (1) when (year of entry), (2) where (institution of entry) and

(3) how (type of entry) this scientist entered the system.

For each researcher, we first define her year of entry as the year[40] in which she published her first paper(s). In addition, we use the address that appears on her first publication(s) to identify her institution of entry. Finally, we define the type of entry by looking at the research environment(s) a scientist was exposed to within her year of entry. We say that a researcher enters with a star if in her first year[41] she publishes a paper with a star scientist. In addition, we say a researcher enters within a particular research environment if she is exposed to one on her first year of entry; this means that a scientist enters with a (1) RG, (2) top-RG, (3) the leader of an RG or (4) the leader of a top-RG if she is exposed to one of these environments in her year of entry. Figures 3.1 and 3.2 depict how the entry space is broken down by type of entry for all the scientists that entered the system between 1984 and 1999. In addition, these figures show in parentheses the productivity by each type of entry. From these figures it can be seen that 71% of the sample can be characterized by one (or more) of our entry definitions. In addition, these figures show that star scientists (with a high rate of citations and productivity) collaborate directly with 18% to 26% of new entrants.

---

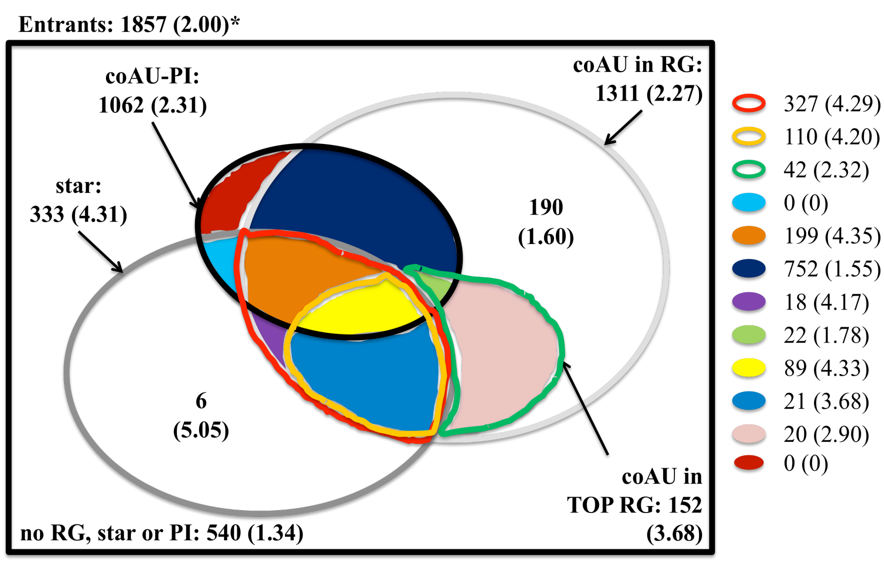
[40] For this analysis we use calendar years to define the year of entry of a scientist.
[41] We use first year (and not first paper) because the time frame of our data is restricted to years; this means that within a particular year we cannot know which paper was published first.



Type of entries: star = Entry with a star; coAU-PI = Entry with a co-author that is a PI of a RG; coAU in RG = Entry with a co-author that is in a RG; coAU in top-RG = Entry with a co-author that is part of a top RG; no RG, star or PI = Entry "alone" or with non of the previous categories.

* Category's productivity within parenthesis.

Figure 3.1. Entry Space Based on Articles. This figure depicts how the entry space is broken down by type of entry for all the scientists that entered the system between 1984 and 1999. In addition these figures show in parenthesis the average productivity (number of articles per year per researcher) by each type of entry.



Type of entries: star = Entry with a star; coAU-PI = Entry with a co-author that is a PI of a RG; coAU in RG = Entry with a co-author that is in a RG; coAU in top-RG = Entry with a co-author that is part of a top RG; no RG, star or PI = Entry "alone" or with non of the previous categories.

* Category's productivity within parenthesis.

Figure 3.2. Entry Space Based on Citations. This figure depicts how the entry space is broken down by type of entry for all the scientists that entered the system between 1984 and 1999. In addition this figure shows in parentheses the average impact (number of citations per year per researcher) for each type of entry.

Finally table 3.6 quantifies the indirect and total contributions[42] each type of star makes to the system. From this table we can see that authors who entered by the hand of highly productive scientists contribute 22% percent of all articles and a quarter of all citations, whereas researchers that entered by the hand of one that has high impact supply 16% of all articles and 17% of all citations. Overall, eminent scientists contribute (directly and indirectly) to two thirds of the system.

Table 3.6. Indirect and Total contribution to the system by Type of Star (One STDev)[1]

Papers			**Citations**		
Indirect Contribution[2]			**Indirect Contribution[2]**		
by author that enters w/star-articles[4]	1414	22%[6]	by author that enters w/star-articles[4]	3129	25%[6]
by author that enters w/star-citations[5]	1077	16%[6]	by author that enters w/star-citations[5]	2189	17%[6]
Total Contribution[3]			Total Contribution[3]		
by star-articles[3]	4154	64%[6]	by star-articles[3]	9497	75%[6]
by star-citations[4]	2899	44%[6]	by star-citations[4]	8452	66%[6]
Total number of papers, 1984-2001	6549		Total number of citations, 1984-2003	12770	

[1] Stars are above the average plus one Standard Deviation.
[2] Indirect Contribution is the total number of published papers by an author that entered with a star, excluding all the ones that were published together; as well as the received citations of those papers.
[3] Total Contribution is the sum of direct (see table 3) and indirect contribution.
[4] Star-articles, stars are defined based on articles per year.
[5] Star-citations, stars are defined based on citations per year.
[6] Percentage from the total number.

3.3.3. Regression Analysis

To quantify the impact different types of entry (i.e. collaboration upon entry with a star scientist or a researcher that is part of an RG or top-RG, or with a PI in an RG or top-RG) have on the productivity of the incoming scientists, we used an ordinary least square (OLS) regression model. We used an OLS (and not a negative binomial) model because our dependent variables (publications per year and citations per year) are continuous (see below for the description of

---

[42] *Indirect contribution* is defined as the total number of published papers by an author that entered with a star, excluding all the ones that were published together; as well as the received citations of those papers. *Direct Contribution* is the total number of published papers by a star, as well as the received citations of those papers. Total Contribution is the sum of both quantities.

these variables). In addition, we employed a logistic regression model to measure the extent to which these new researchers mimic the steps of their mentors and the degree to which different research environments are conducive to becoming a leading scientist.

A key characteristic of our analysis is the use of year and institution fixed effects on our regression models (Mundlak 1978; Hausman and Taylor 1981; Green, 2002). The model controls for otherwise unobserved heterogeneity between year of entry and institution of entry. These controls are important because different institutions will be associated with important heterogeneity in scientific performance and the ability of incoming researchers can vary over time. For example, it is likely that the Physics Department of CINVESTAV, a very well known department, attracts better people for its ranks when compared to a smaller regional university. If this were the case, results of the comparison of productivities for scientists across institutions could be entirely driven by unobserved differences between the institutions, rather than the differences between entry with star and non-star scientists. This could generate misleading results. The model is:

$$Y_{ijt} = \beta_0 + \sum \beta_k x_{kijt} + \varepsilon_{ijt} \tag{1}$$

where

$$\varepsilon_{ijt} = \phi_{ij} + \upsilon_{it} + \omega_{ijt} \tag{2}$$

In this equation, regression coefficients are denoted as β, k indexes the measured independent variables (Xs), i indexes individuals, t indexes time, j indexes institutions, and ε = error terms; ϕ = cross-sectional (institutional) component of error; υ = time-wise component of error; ω = purely random error component; and β_0 = intercept. Y, the dependent variable, is explained in the next section, as well as the different independent ones.

Dependent variables

In order to quantify the increase (or decrease) in productivity and impact that different types of collaborations at the time of entry have on the incoming researchers, two continuous variables were used on the left hand side of the OLS model. The sum of all papers published by a scientist during her tenure,[43] divided by the researcher's tenure, was used as a measure of productivity, whereas the sum of all citations received for each paper (with a three-year window), divided by the scientist's tenure[13], was used to indicate their level of impact. For the logistics model, a dummy variable was used on the left hand side of eq. 1, taking the value of "one" if the new scientist became a star within the studied period and "zero" otherwise. This indicates if this researcher followed the steps of their mentors.

For this work we used the one STDev definition to identify the star scientists and top-RG in our sample. This means that this output is a lower bound (I don't understand your use of the word "bound"; what would it be in Spanish?) for the impact early collaboration with elite scientists and groups have on incoming researchers. In the appendix we show the results for the analysis of the two-STDev definition of star scientists and top-RGs.

Independent variables

Five different dummy variables were used (individually and combined) on the right hand side of eq. 1 to assess the impact different types of entry have on the performance of new researchers and the likelihood of becoming a star scientist; table 3.7 shows these variables.

[43] The *tenure* of a researcher is defined as the total number of years this person was present in the system, from the year she enters the system up to when she leaves it or the year 2001, even if she didn't publish anything in the years in-between.

Table 3.7. Independent variables

Variables	Description
Star (articles or citations)	1 = entry with a star; 0 = otherwise
coAU in RG (articles or citations)	1 = entry with a coauthor that belongs to an RG; 0 = otherwise
coAU in top RG (articles or citations)	1 = entry with a coauthor that belongs to a top RG; 0 = otherwise
coAU-PI in RG (articles or citations)	1 = entry with a coauthor that is a PI in an RG; 0 = otherwise
coAU-PI in top RG (articles or citations)	1 = entry with a coauthor that is a PI in a top RG; 0 = otherwise

3.4. Results of Regression Models

In this section we present the results of the regression models discussed previously. We divide the analysis in two parts. Part one shows the estimates for the increase in productivity and impact for different type of entries, while part two shows the extent to which incoming researchers follow the same path of eminent scientists or if another type of entry is associated with their success, e.g. collaboration with a PI or a co-author that belongs to a highly productive group. The regressions were run first with single variables (e.g. entry by the hand of a star scientist) to assess the individual impact these variables had on the different dependent variables and then with two or more variables (e.g. early collaboration with a star scientist and a co-author that belongs to an RG) to measure the combined effect of these variables on the left hand side of these equations.

3.4.1. Type of Entry and Productivity Impact

In this part we present the results of the OLS regression for papers and citations per year while controlling for different types of entry (tables 3.8 and 3.9). These models show that highly productive scientists have a positive influence on the output of their protégés, boosting their productivity by 27% on average (table 3.8 models AII-01 to AII-04). In addition, the regression models show that entry with a co-author that belongs to a top RG increases the productivity of the new researcher by an average of 18% (table 3.8 models AII-01 and AII-02). Furthermore, entry with the leader of a top RG raises the output of the incoming collaborators by 38%, or 40% more than entering with a star (table 3.8 models AII-03 and AII-04).

Table 3.8. Productivity increase by type of entry, articles per year, 1984-2001

Type of Entry (Std. Err.) [Total Effect*]	**AI-01**	**AI-02**	**AI-03**	**AI-04**	**AI-05**	**AII-01**	**AII-02**	**AII-03**	**AII-04**
Star**, articles	0.296^{c} (0.034) [32%]					0.259^{c} (0.036) [28%]	0.252^{c} (0.037) [27%]	0.244^{c} (0.035) [26%]	0.245^{c} (0.036) [26%]
coAU in RG, articles		0.136^{c} (0.035) [15%]					0.028 (0.037) [NA]		
coAU in top RG**, articles			0.290^{c} (0.050) [31%]			0.168^{c} (0.053) [18%]	0.161^{c} (0.054) [17%]		
coAU-PI of RG, articles				0.098^{c} (0.031) [11%]					-0.005 (0.032) [NA]
coAU-PI of top RG**, articles					0.476^{c} (0.064) [51%]			0.348^{c} (0.066) [37%]	0.350^{c} (0.068) [38%]

* Total Effect = coefficient divided by average.
** Star and top RG defined based on the average plus one standard deviation.
^{a} 10% confidence level, ^{b} 5% confidence level, ^{c} 1% confidence level.
NA coefficient is not significant.

Then we look at the influence of eminent scientists on impact, or number of citations per year, of new researchers. Models CII-01 to CII-04 in table 3.9 show that stars are associated with an increase in the amount of citations received by their advisees by an average of 142%, while other forms of entry have a negligible effect on their citation rate.

Table 3.9. Productivity increase by type of entry, citations per year, 1984-2001

Type of Entry (Std. Err.) [Total Effect*]	CI-01	CI-02	CI-03	CI-04	CI-05	CII-01	CII-02	CII-03	CII-04
Star**, citations	2.811 [c] (0.183) [140%]					2.842 [c] (0.199) [142%]	2.790 [c] (0.201) [139%]	2.866 [c] (0.195) [143%]	2.861 [c] (0.198) [143%]
coAU in RG, citations		0.814 [c] (0.176) [41%]					0.296 [a] (0.173) [15%]		
coAU in top RG**, citations			1.454 [c] (0.273) [73%]			(0.113) (0.283) [NA]	(0.172) (0.285) [NA]		
coAU-PI of RG, citations				0.536 [c] (0.152) [27%]					0.023 (0.151) [NA]
coAU-PI of top RG**, citations					1.515 [c] (0.357) [76%]			(0.302) (0.363) [NA]	(0.309) (0.366) [NA]

* Total Effect = coefficient divided by average.
** Star and top RG defined based on the average plus one standard deviation.
[a] 10% confidence level, [b] 5% confidence level, [c] 1% confidence level.
NA coefficient is not significant.

In addition to independently assessing the effect that highly productive and visible environments have on the productivity and citations of incoming scientists, we can combine both types of environments (publication output and received citations) on a single regression model and measure which milieu has the greatest influence on the publication output and received citations of new researchers. Table 3.10 show that stars with a high citation rate have on average 1.4 more impact on the productivity of new scientists than highly productive ones. In addition, early collaboration with members of a highly productive group (especially the leader of the group) also enhances the productivity of the incoming researcher. In contrast, early collaboration with researchers that belong to a group with high citations rates does not have an impact on the productivity of new scientists. Furthermore, table 3.10 shows that a highly visible star, based on citations, is the only variable that has a significant and positive effect of the level of citations of new researchers.

Table 3.10. Productivity increase by type of entry, articles per year, 1984-2001

Type of Entry (Std. Err.) [Total Effect*]	Articles per year					Citations per year				
	AIII-01	AIII-02	AIII-03	AIII-04	AIII-05	CIII-01	CIII-02	CIII-03	CIII-04	CIII-05
Star**, articles	0.180 [c] (0.040) [19%]		0.153 [c] (0.041) [16%]		0.141 [c] (0.041) [15%]	(0.040) (0.195) [NA]		(0.061) (0.201) [NA]		(0.087) (0.199) [NA]
Star**, citations	0.239 [c] (0.045) [26%]		0.243 [c] (0.047) [26%]		0.229 [c] (0.047) [25%]	2.835 [c] (0.218) [141%]		2.871 [c] (0.230) [143%]		2.900 [c] (0.227) [145%]
coAU in top RG**, articles		0.260 [c] (0.063) [28%]	0.182 [c] (0.063) [20%]				0.376 (0.311) [NA]	0.173 (0.306) [NA]		
coAU in top RG**, citations		0.055 (0.068) [NA]	(0.091) (0.070) [NA]				1.212 [c] (0.338) [61%]	(0.215) (0.340) [NA]		
coAU-PI in top RG**, articles				0.460 [c] (0.078) [50%]	0.364 [c] (0.078) [39%]				0.842 [b] (0.388) [42%]	0.485 (0.380) [NA]
coAU-PI in top RG**, citations				0.032 (0.086) [NA]	(0.112) (0.088) [NA]				0.990 [b] (0.431) [49%]	(0.589) (0.428) [NA]

* Total Effect = coefficient divided by average.
** Star and top RG defined based on the average plus one standard deviation.
[a] 10% confidence level, [b] 5% confidence level, [c] 1% confidence level.
NA coefficient is not significant.

3.4.2. Type of Entry and Likelihood of Also Becoming an Eminent Scientist

In this section we measure the extent to which incoming scientists follow the steps of their mentors. According to the output of the logistics model, a scientist that enters the system by the hand of a highly productive researcher is on average 2.5 times more likely to mimic the success of their mentor and to be regarded as highly productive (table 3.11 models AV-01 to AV-04). In addition, early collaboration with a researcher that belongs to a highly productive group has almost the same effect (table 3.11 models AV-03 and AV-04). These results suggest that a nurturing environment is as important as entry with a highly productive scientist for success, at least in terms of output.

Table 3.11. Likelihood of becoming a Star by type of entry, citations per year, 1984-2001

Type of Entry (Std. Err.)	AIV-01	AIV-02	AIV-03	AIV-04	AIV-05	AV-01	AV-02	AV-03	AV-04
Star*, articles	2.779^{c}					2.317^{c}	2.384^{c}	2.499^{c}	2.591^{c}
	(0.387)					(0.339)	(0.368)	(0.312)	(0.334)
coAU in RG, articles		1.491^{b}					0.893		
		(0.249)					(0.165)		
coAU in top RG*, articles			3.232^{c}			2.205^{c}	2.257^{c}		
			(0.610)			(0.444)	(0.463)		
coAU-PI of RG, articles				1.519^{c}					0.854
				(0.203)					(0.112)
coAU-PI of top RG,* articles					4.849^{c}			2.150^{c}	2.264^{c}
					(1.139)			(0.471)	(0.507)

* Star and top RG defined based on the average plus <u>one</u> standard deviation.
[a] 10% confidence level, [b] 5% confidence level, [c] 1% confidence level.

In terms of visibility (i.e. citations), a new scientist is on average seven times more likely of becoming a star if he or she enters the system by the hand of star (table 3.12). However, early co-authorship with a researcher that belongs to highly cited groups has in the best case a small effect (at least when compared to entry with a star) on the chances of the new scientist (table 3.12 model CV-02) or no effect at all (table 3.12 model CV-03 and CV-04). These results suggest that the protégées of highly cited scientists will mimic the success of their mentors.

Table 3.12. Likelihood of becoming a Star by type of entry, citations per year, 1984-2001

Type of Entry (Std. Err.)	CIV-01	CIV-02	CIV-03	CIV-04	CIV-05	CV-01	CV-02	CV-03	CV-04
Star*, citations	6.904^{c}					8.240^{c}	6.011^{c}	8.713^{c}	7.084^{c}
	(1.141)					(1.226)	(1.097)	(1.276)	(1.302)
coAU in RG, citations		1.865^{c}					0.974		
		(0.383)					(0.221)		
coAU in top RG*, citations			4.332^{c}			1.229	1.785^{b}		
			(0.927)			(0.240)	(0.425)		
coAU-PI of RG, citations				1.444^{b}					0.759
				(0.226)					(0.138)
coAU-PI of top RG*, citations					4.141^{c}			1.004	1.511
					(1.150)			(0.242)	(0.458)

* Star and top RG defined based on the average plus <u>one</u> standard deviation.
[a] 10% confidence level, [b] 5% confidence level, [c] 1% confidence level.

In terms of the impact that combined environments (i.e. different types of eminent scientist – articles vs. citations – or other forms of early collaborations) have on the likelihood of replicating success, table 3.13 shows that a new scientist is more likely to become a star (based on articles) if he or she enters the system with a highly cited researcher instead of a highly productive one (models AVI-01, AVI-03 and AVI-05). In addition, early collaboration with researchers belonging to highly productive groups also has a positive and significant effect on the chances of the new scientist (models AVI-02 to AVI-05). Furthermore, models CVI-01, CVI-03 and CVI-05 of table 3.13 show that entry with a highly cited scientist has the highest effect, compared to any other form of entry, on the likelihood of new researchers to replicate the same success. Models CVI-02 and CVI-03 show that early collaboration with a co-author that belongs to a top RG, based on citations, also has a positive effect of the chances of becoming a highly visible star.

Table 3.13. Likelihood of becoming a Star by type of entry, 1984-2001

Type of Entry (Std. Err.)	**Based on articles per year**					**Based on citations per year**				
	AVI-01	**AVI-02**	**AVI-03**	**AVI-04**	**AVI-05**	**CVI-01**	**CVI-02**	**CVI-03**	**CVI-04**	**CVI-05**
Star*,	0.593 ^{c}		1.607 ^{c}		1.556 ^{b}	0.560 ^{b}		0.600 ^{b}		0.566 ^{b}
articles	(0.169)		(0.279)		(0.270)	(0.126)		(0.139)		(0.130)
Star*,	0.812 ^{c}		2.104 ^{c}		2.159 ^{c}	9.847 ^{c}		8.184 ^{c}		9.218 ^{c}
citations	(0.176)		(0.390)		(0.397)	(2.161)		(1.842)		(2.077)
coAU in top		2.580 ^{c}	2.005 ^{c}				0.537 ^{b}	0.584		
RG*, articles		(0.638)	(0.513)				(0.170)	(0.192)		
coAU in top		1.472	0.961				6.466 ^{c}	2.571 ^{c}		
RG*, citations		(0.397)	(0.270)				(1.924)	(0.806)		
coAU-PI in top				4.195 ^{c}	3.083 ^{c}				0.944	0.892
RG*, articles				(1.228)	(0.941)				(0.355)	(0.335)
coAU-PI in top				1.327	0.812				4.289 ^{c}	1.521
RG*, citations				(0.450)	(0.286)				(1.542)	(0.556)

* Star and top RG defined based on the average plus one standard deviation.
^{a} 10% confidence level, ^{b} 5% confidence level, ^{c} 1% confidence level

Overall, these results show that highly productive and cited researchers have a positive and significant effect on the productivity of early career scientists and their ability to replicate the same success. In addition, the outcome of these models suggests that early collaboration with scientists that are embedded in productive environments also enhances the productivity of new researchers.

3.5. Conclusion and Policy Implications

In the last decades, ST&I have been seen as a major source of economic growth. To be able to leverage ST&I, government officials and policy makers around the world have developed policies to develop their systems. At the core of these initiatives are strategies to expand the scientific base and to generate, attract and retain highly talented scholars. The main rationale behind these developments is that key scientists play a vital role in the growth of this system.

In the past, several authors have studied the impact eminent scientists have on the development of the research system, the institutions they belong to and the scientists they work with. In addition, scholars have looked at the conditions that foster highly productive and talented scholars. This paper contributes to this body of knowledge by assessing the impact key scientists have on the development of the system. In particular, this work quantifies the effect that star scientists have, by themselves and within the context of different research environments, on the productivity and impact of young faculty, as well as on the likelihood of also becoming a leading personality in science.

Our analysis confirms our expectations and previous results that eminent scientist have a prime role in the development of a scientific system, especially within the context of an emerging economy like Mexico. In particular, in terms of productivity and visibility, this work shows that between 1984 and 2001 the elite group of physicists in Mexico (approximately 10% of all scientists working in physics and related fields) authored 42% of all publications and received 50% of all citations and bred 18% to 26% of new entrants.

In addition, our work shows that scientists that entered the system by the hand of a highly productive researcher increased their productivity on average by 28%, and those that did so by the hand of a highly visible scientist received on average 141% more citations, vis-à-vis scholars that did not publish their first manuscripts with an eminent scientist. Furthermore, incoming scientists also had an additional boost in their productivity if they were exposed early on in their careers to the appropriate research environment. For example, young faculty increased their publication rate by 18% if they had an early collaboration with a scientist belonging to a highly productive research group, and 38% if this collaborator was the leader of the group. But these environments did not have any effect on the citation rate of new faculty. In summary, key scientists have a positive and significant effect on the productivity and visibility of young faculty, but nurturing environments only impact their productivity.

In terms of mimicking success, this work shows that scientists working at the beginning of their careers with eminent researchers tend to replicate the success of their mentors. In particular, scholars that enter the system by the hand of a highly productive researcher were on average 2.5 more likely to also become a star, when compared to the ones that did their initial work with non-star scientists; and early collaboration with highly visible researchers increased 7.4 the chances of a new scientists mimicking the success of their mentors. In addition, early collaboration with scholars belonging to a highly productive group and the leaders of such ensembles also had on average an additional impact of 2.2 on the probability of someone becoming a leading personality in science; whereas early collaboration with a co-author belonging to a highly visible group was the only environment that had an additional effect on someone's likelihood of becoming a highly visible scientist; in this case by 1.8 percent.

These results have important consequences for policy-making in science, technology and innovation systems. They tell us that eminent scientists have a primary role in the growth of these systems and the development and productivity of young faculty. In addition, they show that nurturing environments play an additional role in the construction of these systems. This means that if a country or region wants to improve or become the leader in a certain area of knowledge it should focus on attracting and retaining the best and the brightest, and creating around these key figures appropriate collaborative environments so that research and researchers can flourish.

Chapter 4. Learning and Opportunities in Collaborative Research Environments

4.1. Introduction

In the last decades, science, technology and innovation (ST&I) have played a major role in fostering economic development and improving the welfare of society (Wagner et. al, 2001). Leydesdorff and Etzkowitz (2001) describe this evolution as a Triple Helix, which reflects an important transformation in the relationships between the university, industry, and government, with academia playing an enhanced role in innovation (Etzkowitz & Leydesdorff, 2000) and engaging in new functions, such as promoting firm-formation (Lissenburgh & Harding, 2000) and regional development (Etzkowitz, 2001).

A major and growing force (deB. Beaver, 2001; Wagner and Leydesdorff, 2005) emerging alongside the development of the scientific endeavor is the ability of individual scientists to initiate and sustain productive partnerships with other scholars and foster collaborative research environments within universities (as shown in chapter three of this thesis) and with industry (Gibbons et al., 1994). To promote cooperation within academia and with firms, a variety of nations have adopted policies to promote collaboration within their ST&I system, across sectors and with other nations (Luukkonena, 2001). This has motivated an important stream of research on scientific collaboration. According to Bukvova (2010), studies have primarily focused on five main areas: (a) defining what is research collaboration; (b) problems with measuring this phenomenon; (c) understanding why researchers collaborate; (d) developing explanatory approaches to research collaborations and, in recent years, (e) understanding the role Information and Communication Technologies (ICT) plays in promoting this type of activities in science. There has been a particular effort to uncover the factors that promote and hinder scientific collaboration, as well as to understand the role that research collaboration plays in the development of science (Bukvova's, 2010). Similarly, studies have tried to assess the costs, benefits and opportunities these schemes produce for individual researchers throughout their professional career (Katz and Martin, 1997; deB. Beaver, 2001).

Past research has shown the importance that different research environments, and the interactions that happen within them, have on the development of researchers. According to Hemlin et al., (2009), the degree to which an "individual's creative potential is expressed depends considerably on the environment in which that individual works. To understand scientific and technological creativity, one needs to analyze the interactions between individuals or groups and their environment." Others have shown how collaboration, seen as social interaction, is a key condition for the emergence of creativity in science (Zuckerman, 1987; Laudel, 2001). Furthermore, scholars have assessed the impact these environments can have on the professional career of scientists. For example, Crane (1965) shows that the caliber of the institution has a positive impact on the productivity and prestige of its scientists, while Oettl (2009) notes that productive and collegial milieus have a higher impact on the productivity of researchers than only productive or cooperative setups.

Some authors have also considered the extent to which ensembles of scientists provide a nurturing environment where researchers, in particular younger ones, can flourish. Bozeman and Corley (2004) emphasize the importance of collaboration on the development of the scientific and technical human capital[44] of researchers, especially when a senior scientist works with a

[44] Bozeman et al (2001) defines "Scientific and technical human capital" (S&T human capital) as "the sum of researchers' professional network ties and their technical skills and resources."

junior one and the former acts as her mentor. According to these authors, under the right circumstances, a graduate student or post-doctoral researcher can gain, "not only enhanced S&T knowledge, but craft skills, know- how, the ability to structure and plan research and, of course, increase contacts with other scientists, industry, and funding agents." Chapter 3 complements this perspective by showing that leading researchers have a positive effect on the performance of young researchers and on the likelihood that they also become star scientists. Similarly, they find that early collaboration with a highly productive research group and the leader of this group also contributes to the future productivity of new scientists.

Despite these advances in the assessment of the effect of research collaboration on the scientific endeavor, there are still many gaps in our understanding of the phenomena. First, research on scientific collaboration has primarily focused on understanding the role it plays throughout the career of the scientist, overlooking the contributions this activity provides at different stages of the researcher's professional life, especially at the beginning of their career. In addition, previous work has typically considered collaboration in general terms, or focused on particular research environments (like only collaborating with eminent scientists), failing to notice the combined effect that different environments can have on scientists. Publishing with a highly productive single researcher or being part of a highly productive research group can mean different things and have diverse impacts on the career of a scientist. Finally, with few exceptions (Duque et al., 2005; Wagner, 2008; chapters 2 and 3), work on research collaboration tends to focus on developed nations. This work tries to bridge some of the existing gaps in the literature by assessing the impact that different forms of collaborations have on the future career of new scientists. In particular, we asked a group of researchers in an emerging economy about the opportunities they received and what they learned from their initial relationships with a variety of research settings.

The work is divided into four sections. First we lay out the purpose of this study and its research questions. Next, we define the methodology for this work, explain the data and define some key concepts. Third, we analyze the data and provide results. Finally we end the paper with policy implications and concluding remarks.

4.2. Literature Review and Research Questions

Previous studies on research collaboration, and the collaborative environment that emerges from these interactions, have mostly focused on understanding the role scientific cooperation plays in the development of science (deB. Beaver and Rosen, 1978, 1979a, 1979b; Wagner, 2008; Bukvova, 2010), uncovering the factors that promote this activity (Katz and Martin, 1997; deB. Beaver, 2001), quantifying the impact these interactions have on the performance of scientists (Azoulay et al., 2007; Ottle, 2009; Waldinger, 2010), or assessing the effect that different collaborative setups have on the professional career of researchers (Reskin, 1979; Ottle, 2009; chapter 3). In this study, we also focus on the impact that scientific collaboration at the individual level, as well as collaborative research environments, have in the professional development of scientists. However, instead of assessing the effect that ongoing or recent research collaborations with their peers have on scientists and their performance, we will consider the qualitative impact that a variety of interactions at the beginning of their professional career have on their behavior. Furthermore, we assess the opportunities that these early relationships open, as well as what researchers learn within these settings.

4.2.1 Research Collaboration

The surroundings in which research is conducted have been seen to influence the productivity and impact of scientists. For example, working with highly accomplished researchers (like Nobel laureates) has a significant positive effect on productivity and impact (Oettl, 2009: Azoulay et al., 2010; Waldinger, 2010), as well as the chances of also becoming highly regarded, i.e. also being recognized with a prize (Zuckerman, 1967; Ham and Weinberg, 2007). The performance of a scientist can also be influenced by the caliber of the institution she belongs to (Crane, 1965; Allison and Long, 1990) and its organizational context, such as industry or academia (Reskin, 1977; Long and McGinnis, 1981).

Given this relevance, it is not surprising that the scientific community has dedicated great effort to the understanding of the factors that promote, enhance and hinder collaboration in science. These factors include individual characteristics, such as particular personalities being suited for collaborative work and leadership (Stokols et al., 2008); group attributes, such as size (Rigby, 2009), and ability to coordinate (Cummings and Kiesler, 2007), communicate (Stokols et al., 2008) and deal with differences (Jeffrey, 2003; Bammer 2008); institutional features, such as academic culture (Sorensen, 2003) or granting scientific credit (Kennedy, 2003; Birnholtz, 2008); and National/International Science Policy, such as funding (Defazio et al., 2009) or national security (Dias et al., 2010).

A complementary perspective has specifically considered how the research environment in which an individual is embedded matters for his creativity[45] and performance. Research shows that to understand creative processes in science and the production of knowledge, one needs to

---

[45] According to Hemlin et al., (2009) creativity refers to the "generation of a product that is not only novel and imaginative but also useful and of good quality."

consider the environment (Amabile and Gryskiewicz, 1989; Witt and Beorkrem, 1989; Woodman et al., 1993; Hunter et al., 2007) and the interactions that researchers have within these setups at different levels of the organization (Hemlin et al., 2009).

If collaboration does indeed matter, one may think that it would be particularly significant when considering young scientists and their initial steps into the scientific endeavor. For example, Bozeman and Corley's (2004) work on scientific and technical human capital has shown the importance that research collaboration can have on the development of capabilities among junior scientists, especially when they work under the supervision of a senior researcher who acts as a mentor. According to these authors, under the right circumstances, a graduate student or postdoctoral researcher can gain, "not only enhanced S&T knowledge, but craft skills, know-how, the ability to structure and plan research and, of course, increase contacts with other scientists, industry, and funding agents" (Bozeman and Corley, 2004). They can also learn how to conduct research and get published (Reskin, 1979; Judge et al., 2004), critical tools for the budding researcher.

When considering the surroundings in which research is conducted, academic stars, such as Nobel laureates, have been found to be of particular relevance for the productivity and impact of other scientists. This group of scholars is important because they contribute to the system with a large amount of papers (Lotka, 1926; Pirce, 1963; Zucker, Darby, and Brewer, 1998) and citations (Hagstrom, 1968; Cole, 1970; Cole and Cole, 1972; Allison and Stewart, 1974). In addition, this small percentage of researchers has a significant positive effect on the productivity and impact of its collaborators (Oettl, 2009: Azoulay et al., 2010; Waldinger, 2010) and their chances of also becoming highly regarded, i.e. also being recognized with a prize (Zuckerman, 1967; Ham and Weinberg, 2007).

While interacting with a prominent researcher can have a profound effect on a scientist, it is clearly not the only dimension that could have an impact on their professional career. The surroundings in which research is conducted also influence the productivity and impact of scientists, in particular graduate students and postdoctoral fellows (Fox, 1983; Fox and Mohapatra 2007; Louis et al. 2007). For example, the performance of a scientist can be influenced by the caliber of the institution she belongs to (Crane, 1965; Allison and Long, 1990) and its organizational context, such as industry or academia (Reskin, 1977; Long and McGinnis, 1981). In addition, work at the department level has also found that scholars publish more when they are surrounded by productive peers (Braxton, 1983) and research-oriented coworkers (Bland and Ruffin, 1992).

Leadership within research organizations or groups is a variable that affects different factors, which in turn may also influence research productivity (Bland and Ruffin, 1992). Team leaders are there to "influence member's knowledge and values, facilitate contact and networks, attract other competent researchers, and help colleagues who are blocked or stopped in their researcher efforts" (Bland and Ruffin, (1992) citing Dill (1985, 1986)), and they are crucial for enhancing productivity and boosting the morale in turbulent times (Ramsdena, 1998). This suggests that it would be important to consider the impact that different research environments may have at the beginning of a scientist's professional career, which is the focus of this study.

Many authors have looked at the benefits and costs of scientific collaborations to individual researchers throughout their careers. Cooperation in science can produce many advantages, including access to expertise, funding and resources (like instrumentation and datasets), ability to exchange ideas (especially across disciplines), learning new skills, pooling expertise for complex problems, prestige, and, in some cases, fun and pleasure (Katz & Martin, 1997; deB Beaver, 2001; Bukvova, 2010). However, collaboration is not risk-free activity. In some instances, this collaborative endeavor can lead to added financial costs (from travel or relocation), an increase in the amount of time spent on research, or greater bureaucratic and managerial costs; it can also be difficult to reconcile organizational and cultural differences between collaborators (Katz and Martin, 1997). In addition, it can amplify coordination costs (Cummings and Kiesler, 2007) and in some instances it can be difficult to assign credit to the participants (Wray, 2006). Table 4.1 presents a summary of the cost and benefits of research collaboration.

Table 4.1. Impact of Research Collaboration

Positive effects	Negative effects
Access to expertise	Financial costs (from travel or relocation)
Access to resources	Increase in the amount of time spent on research
Exchange of ideas	Increase in bureaucratic and managerial costs
Pooling expertise for complex problems	Difficulty in reconciling organizational and cultural differences
Learning new skills	Amplification of the coordination costs
Higher productivity	Assigning of scientific credit
Access to funding	
Prestige	
Political factors	
Personal factors	
Fun and pleasure	

Based on Katz and Martin, 1997; deB Beaver, 2001; Bukvova (2010).

While existing research has documented a variety of important factors associated with scientific collaboration and their impact on performance, much less has been considered in terms of how scientists value and leverage the various types of collaborations. In addition, there is virtually no work considering how young scientists value and leverage collaborative environments. In this research, we look at the benefits that research collaboration produce on new researchers. In particular, we assess the opportunities these interactions open to incoming scientists, as well as what they get in terms of learning.

4.3. Methodology

For this study we used a descriptive approach because we were interested in a rich, detailed picture of the impact that early collaborations, and the different research environments that emerge from these interactions, have on the development of new researchers, as well as the opportunities they create and the knowledge that is shared within these setups. In particular we used an online survey to collect information about the characteristics, actions, or opinions of a large group of people (Pinsonneault and Kraemer, 1993p. 77), elicit their attitudes (McIntyre, 1999, p. 75) and examine impacts in detail (Salant & Dillman, 1994, p. 2). An online questionnaire is a particularly useful instrument because it allowed us to reach a large number of scientists who provided detailed information about their early collaborations in science. The data is self-reported, which raises concerns with respect to response accuracy and selection bias. However, we believe that the benefits of such an online survey outweigh the drawbacks (for a discussion, see Bertrand and Mullainathan, 2001).

4.3.1. Definitions

For this analysis, we classify all the researchers in our sample along several dimensions.

Star and non-star scientists

First, we characterize the researchers in our study as star (or eminent) and non-star scientists based on how they perceived their own performance and the performance of their close collaborators vis-à-vis their peers along two dimensions: productivity (measured in terms of papers) and impact (calculated in terms of citations). This means that a researcher is considered a star scientist if she considers that her peers see her as playing a prominent role in academia when compared to other significant researchers in the field of study.

Great effort has been dedicated by the research collaboration literature in answering the questions of what is research collaboration and how it can be measured. According to Katz and Martin (1997), research collaboration is not easy to define because it is "largely a matter of social convention among scientists" that varies "across institutions, fields, sectors and countries" and is not invariant over time. Cooperation in science can manifest itself through different forms of activity and only a few of these can be captured through co-authorship (Laudal, 2002). In spite of this, prior research has used different dimensions to describe it; for example, the professional background and institutional affiliation of the participants, their disciplinary focus, their geographical location and the organizational level where these interactions occur (Amabile et al., 2001; Sonnenwald, 2007). This means that research collaboration is a social process that takes place between individuals who primarily are researchers and belong to one or more institutions from one or more regions of the world. These interactions happen within and across fields of knowledge and, in some same cases, this cooperation happens at different organizational levels, such as departments or institutions (Bukvova, 2010). Some authors have tried to define research collaboration explicitly. For example, Laudel (1999, p.32; 2002) defines research collaboration as a "system of research activities by several actors related in a functional way and coordinated to attain a research goal corresponding with these actors' research goals or interests."

For this study, we define research collaboration based on co-authorship of book and articles (ISI and non-ISI papers), i.e., two or more researchers are collaborators if they have published a manuscript together. In addition, we define initial research collaboration based on the first five manuscripts[46] a researcher published in her scientific career. While this is an objective characterization, we also recognize that such an approach has a variety of limitations identified by the literature in terms of characterizing cooperation in science (Katz and Martin; 1997). The survey presents this logic and asks the respondents to answer a series of questions based on the experience of the researcher in the context of these early publications.

Research Groups and Principal investigator

One of the key environments considered in our work is one of the Research Group (RG). This is defined as a group of people that collaborate repetitively in scientific research and publish the results of these activities in articles or books. In addition, we also characterize an individual as the Principal Investigator of a research group (PI-RG) if this person is the scientific leader of the group. Furthermore, following the same logic of the star scientist, an RG (principal investigator) is considered to be a highly productive group (top-PI) if the ensemble (individual) is seen as having played a leading role in academia when compared to other relevant groups (scientists).

Research environments

Based on the previous definitions we define thirteen research environments based on different publishing contexts of the first five papers (or fewer, if the researcher had a lower number of publications): (1) publish alone; co-author these articles with (2) one, (3) two to six, or (4) seven or more co-authors; work with a (5) highly productive scientist or one that has a (6) high impact; publish these manuscripts within a (7) RG or with a group that is (8) highly

[46] If a scientist had less than five publications we used thiese manuscripts as her initial collaborations.

productive or has a (9) high impact; co-author a paper with the (10) PI of an RG, a (11) highly productive PI and a (12) PI with high impact; or (13) other form of collaboration.

Factors promoting initial collaborations

We define four ways to start collaborating with other researchers at the beginning of the career of a scientist: formal collaboration, when a scientist is invited to collaborate after participating in a competitive process at a research institution/program and being admitted to it (i.e., admission to a PhD program or post-doctoral position); informal collaboration, when an author is invited to collaborate without such competitive process; collaboration sought by the researcher, when this person actively sought a particular collaboration herself; and other types of collaborations includes relationships that are not included in those previously cited.

Entry and tenure

Finally, in this study we say that a researcher entered the scientific system in a particular year when she published her first article or book. In addition, we define the tenure of a researcher based on the number of years this person has been in the system or since she published her first manuscript.

4.3.2. Data Collection

To fully understand the effects that research collaborations, and the different environments that emerge from these, have on new researchers, we developed an online questionnaire (appendix 1 has a sample of this instrument). In addition to socio-demographic questions such as gender and academic rank, we asked respondents about their initial interactions in science, focusing on the following topics: (1) how the researcher started his/her early collaborations in science, (2) the impact of different environments, and the interactions within these settings, on incoming scientists, (3) the opportunities these relationships presented, and (4) what the new researchers got out of these partnerships.

The survey was designed to last approximately 20 minutes, and it was uploaded to Survey Monkey[47] and customized for full usability on this website. Before administering this instrument, we tested it with two researchers from the Physics Department at CINVESTAV and incorporated some of their comments in the final version[48]. All the scientists that belong to the field of physics and related areas (like applied physics, optics and material science among others) in Mexico that were part of CONACYT's National Research System by the end of 2010 were invited to participate. We collected all the answers in March of 2011.

Before continuing with this work it is appropriate to give a small description of Mexico's National Research System, or SNI (acronym in Spanish). The SNI System was created in the1980s by the Mexican Government to recognize the scientific and technological contribution of researchers in this country. The recognition is based on peer review evaluations and grants the appointment of National Researcher. This system has four levels, which are based on performance: candidate (which usually is the entry level to the system) and levels one, two and three, where the last level is the highest recognition within this system. Parallel to the appointment, the researcher receives an economic incentive based on the tier she belongs to (CONACYT, 2012).

---

[47] http://*www.**surveymonkey**.com.*

[48] This survey was approved by the Carnegie Mellon University Institutional Review Board (www.cmu.edu/osp/regulatory-compliance/human-subjects.html).

4.4. Analysis and Results

4.4.1. Summary Statistics

In this section we describe our sample in terms of tenure, eminence and performance. For such purpose, we distinguish between two forms of eminence: one based on our definition of star scientist, noted previously; and the second based on the different levels of the National Research System of CONACYT. Table 4.2 shows the size of our sample, the number of physicists that answered our survey and the response rate. From this table it can be seen that almost 20% of physicists in Mexico answered our survey.

Table 4.2. Sample space and response space

	Number of Researchers
Sample Space – Physics and related areas	1358*
Number of people that:	
- Started the survey	346
- Completed the survey	263
Response rate:	19.4%

* This number corresponds to all researchers that were part of SNI at the beginning of 2011 in Physics and related areas, before we applied the online survey (CONACYT, 2011).

Figure 4.1 shows the distribution of researchers by tenure and gender. From this figure it can be seen that our sample is relatively "young" in tenure, since 45% of it has less than 10 years in the system. In addition, there is on average a 3:1 male to female ratio in our sample. According to recent data from CONACYT, Mexico has 1358 researchers in the field of Physics and related areas, among which 195 (or 14%) are women (CONACYT-SIICYT, 2011).

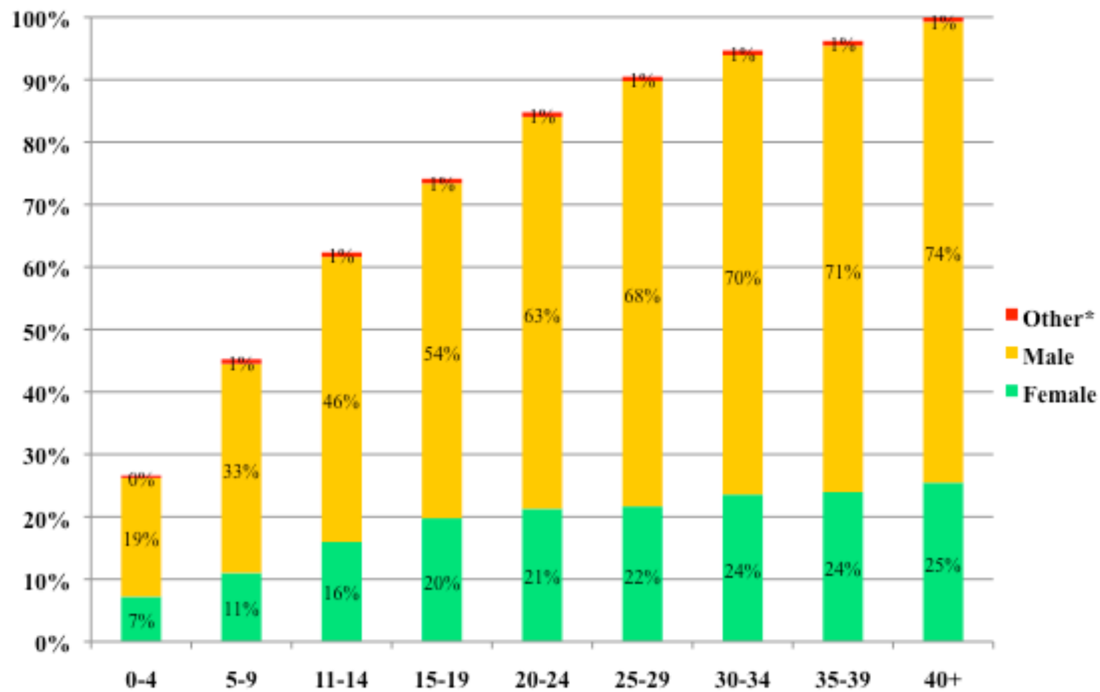


Figure 4.1. Distribution of Researchers by Tenure and Gender.

Table 4.3 reports the distribution of researchers by SNI level (real and sample) and self-reported eminence. We can see that 21% of our sample are candidates, 43% belong to SNI level 1, 21% to level 2 and 12% to level 3. In addition, it can be seen that the candidate category is slightly overrepresented, although the distribution of researchers appears to be the same. Furthermore, this table shows that our sample has a high self-esteem, since 32% think they are on average more productive than their peers and 29% think they have a higher impact. But self-reported eminence is correlated with the SNI level, which is positive.

Table 4.3. Distribution of Researchers by SNI level and self-reported eminence

	SNI Distribution			Researcher sees herself as a star-scientist									
				Articles					Citations				
	Real[(1)]	Sample		No		Yes		NA	No		Yes		NA
	%	#	%[(2)]	#	%[(3)]	#	%[(3)]		#	%[(3)]	#	%[(3)]	
None	0%	5	2%	4	80%	1	20%		3	60%	0	0%	2
Candidate	16%	55	21%	49	89%	6	11%		51	93%	2	4%	2
Level 1	45%	114	43%	84	74%	29	25%	1	79	69%	24	21%	11
Level 2	23%	56	21%	32	57%	24	43%		29	52%	27	48%	
Level 3	16%	32	12%	8	25%	23	72%	1	7	22%	24	75%	1
Blank		1	0%					1					1
Number, %[(4)]		263	100%	177	67%	83	32%	3	169	64%	77	29%	17

[(1)] Percentage of the total number of SNI researchers in Physics and related areas (CONACYT-SIICYT, 2011).
[(2)] Percentage within SNI Level of our sample.
[(3)] Percentage from total (263).
[(4)] Percentage across categories.
NA = No Answer.

4.4.2. Entry into the system and factors promoting initial collaborations

Table 4.4 illustrates how our sample entered the system. These data were aggregated in groups of three columns showing (from left to right) the number of researchers that entered with other researchers, a Research Group, or the leader (w/ Principal Investigator) of an RG. In addition, the last two columns of each triplet show the number of researchers that were exposed early on in their career to a top-performing environment, i.e, an eminent scientist (w/ Star or Top PI) or top RG. From this table it can be seen that, in general, these scientists started collaborating early on in their careers, with 95% of them working with other researchers on their first manuscripts. Moreover, 3 out of 4 did it with an RG and 67% collaborated directly with the leader of a group. In addition, 44% of our sample were exposed to a top-performing environment – such as star scientist, a top-RG or a top-PI. Furthermore, this exposure happened at a higher degree for milieus with high productivity than those with a high citation rate. For example, 70% of our sample reported that they collaborated at the beginning of their career with a highly productive scientist, while only 58% did so with a high impact researcher.

Table 4.4. Entry into the system

	Type of entry								
	w/ another researcher			w/ Research Group (RG)			w/ Principal Investigator (PI)		
		w/ Star			Top RG			Top PI	
N=263		**Art.**	**Cit.**		**Art.**	**Cit.**		**Art.**	**Cit.**
Yes	249	183	152	201	139	118	177	146	117
Percentage									
Total[1]	95%	70%	58%	76%	53%	45%	67%	56%	44%
Within category[2]		73%	61%		69%	59%		82%	66%
No	13	65	76	48	60	65	22	31	47
No answer	1	15	35	14	64	80	64	86	99

[1] Total: Percentage based on the total number, N=263
[2] Within category: Percentage based on the total number within category: N_1=249, N_2=201 and N_3=177.

Table 4.5 shows how the scientists in our sample initiated their first collaborations. It provides the frequency rate (based on the total sample) of these initial publications by research environment and collaboration type. The table suggests that the mechanisms by which collaboration is initiated relate differently to the various collaborative environments that researchers can experience. First, structured competitive processes, like applying to a Ph.D. or postdoctoral program, tend to be associated with greater levels of collaboration with a (formal) research group, or its PI. In contrast, informal processes are more relevant when linking with an informal group of scientists (two to six coauthors). In addition, it seems that new researchers will actively seek to collaborate with specific researchers or the leaders of research groups.

Table 4.5. Factors promoting initial collaborations

Research Environment* (percentage relative to category)	Type of Publishing Initiation: Formal	Informal	Search by researcher
a. Alone	19%	18%	12%
b. Only one coauthor	31%	35%	21%(2)
c. Two to six coauthors	40%	56%(1)	21%(2)
d. Seven or more coauthors	17%	22%	8%
e. Star scientist – Art.	37%(3)	37%	17%
f. Star scientist – Cit.	32%	35%	16%
g. RG	49%(1)	43%(2)	17%
h. Top RG – Art.	37%	31%	14%
i. Top RG – Cit.	34%	30%	13%
j. PI	44%(2)	39%(3)	22%(1)
k. Top PI – Art.	35%	36%	20%
l. Top PI – Cit.	32%	33%	17%

* First five or fewer papers published within contexts (a) to (h).
NOTE: There can be double counting across research environments.
(X) Indicates the collaboration startup with the highest percentage within collaboration categories.

4.4.3. Impact, opportunities and learning

As previously noted, the various avenues for research collaborations are expected to have diverse benefits and costs to scientists throughout their professional career. In this work we sought to better understand the impact that the different research environments that young scientists were exposed to had on their career. In particular, we assessed the impact that different forms of entry had on the scientists' self-evaluation of eminence. In addition, we look at the opportunities these setting opened to them, as well as the knowledge they gained from these environments.

Table 4.6 shows the impact that different forms of entry had on the likelihood of someone perceiving herself to be a top-scientist. The table suggests that early collaboration with a top-research environment has a positive effect on the scientist's perception of her own eminence; for example, scientists that cooperated with a top researcher at the beginning of their career had about a 50% chance of seeing themselves as a prominent scholar; whereas the ones that did not collaborate with this type of environment had a little bit more than a 25% chance of having an equivalent perception. This asymmetry is also prevalent in other environments. Moreover, the influence of any of the three critical collaborative dimensions on the likelihood that the young scientist becomes a star is equivalent. The data suggest that the odds that a budding scientist reaches eminence increases by 1.6 to 1.7 when exposed to any of these critical collaborative environments[49].

Table 4.6. Entry into the system and self-evaluation of eminence

Type of entry	Odds that a researcher sees herself as a star scientist			
	Articles		Citations	
	Odds	Odds ratio	Odds	Odds ratio
Entry with a star (N=249)				
Based on Articles	**74:175**	**42%**		
Yes	60:123	49%		
No	14:52	27%		
Based on Citations			**70:179**	**39%**
Yes			50:102	49%
No			20:77	26%
Entry with a top RG (N=201)				
Based on Articles	**59:142**	**42%**		
Yes	45:94	48%		
No	14:48	29%		
Based on Citations			**55:146**	**38%**
Yes			37:81	46%
No			18:65	28%
Entry with a top PI (N=177)				
Based on Articles	**52:125**	**42%**		
Yes	47:99	47%		
No	5:26	19%		
Based on Citations			**48:116**	**41%**
Yes			38:79	48%
No			10:37	27%
Total	**83:180**	**46%**	**77:186**	**41%**

[49] It's greater for entering with a Top PI based on articles. Yet the number of observations is smaller for one of the categories, which might justify the unbalance.

Table 4.7 provides the response rate for the open question about "which research environment had the highest impact on your scientific career?[50]." It considers three views of the population: (1) all the researchers in the sample; (2) a breakdown according to self-reported eminence of productivity and impact, as well as a cutoff according to (3) SNI level at the time that the questionnaire was implemented. For the first two categories there seems to be a consensus (with slight ordering differences) that research environments characterized by only one co-author, two or more co-authors, as well as an RG, had the highest impact on the career of those in our sample. In addition, the first two categories were also highly regarded by all SNI levels as being important. It is interesting to note that for the first two categories, as well as the first two SNI levels, early collaboration with star scientists and top PIs, do not seem to matter too much for this group of researchers. In contrast, there are important differences for scientists at level 3, where collaboration with star scientists and top PI were seen as having an important role in the career of this group of people. In fact, the data implies that the higher the SNI level the higher the important these two environments have on the development of researchers, for example, the importance of top PIs ranged from 7% at the candidate level up to 22% at level 3, the highest possible in SNI, whereas prominent scientists ranged form 11% up to 22%. In addition, these results also suggest that top scientists, such as those that have reached the top tier, saw a greater benefit in their development if they collaborated with only one researcher, in particular if this scholar was also an eminent one.

[50] For this question we received different types of answers, ranging from letters which made reference to the predefined researcher environments *a* to *l*, defined previously, to small paragraphs describing the setting that had the highest impact. The former set of answers were analyzed and converted to fit one of the following publishing milieus: a) alone, b) only one co-author, c) two or more co-authors, d) star scientists, e) RG, f) top RG, g) PI and h) top PI. In some cases, there was a 1-to-1 relationship between the answer and a setting, but in others an answer could fit more than one environment, in which case we included all the milieus that were relevant.

Table 4.7. Perceived Impact* of Research Environments upon entry and Eminence

Research Environment** (percentage within category)	All	Researcher sees herself as a star: Based on Art.		Researcher sees herself as a star: Based on Cit.		SNI Level				
		Yes	No	Yes	No	None	Candidate	Level 1	Level 2	Level 3
a. Alone	12%	14%	11%	13%	11%	0%	4%	10%	23%	16%
b. Only one coauthor	26%(3)	30%(2)	23%(2)	31%(2)	22%(3)	33%	15%	27%(3)	30%(1)	31%(1)
c. Two or more coauthors	35%(1)	35%(1)	34%(1)	34%(1)	33%(1)	17%	35%(1)	39%(2)	29%(2)	31%(1)
d. Star Scientists	17%	14%	18%	19%	14%	0%	11%	17%	21%	22%(2)
e. RG	32%(2)	27%(3)	34%(1)	23%(3)	32%(2)	67%	27%(2)	40%(1)	27%(3)	9%
f. Top RG	12%	11%	12%	13%	11%	0%	15%	10%	13%	16%
g. PI	20%	20%	19%	22%	18%	17%	22%(3)	23%	16%	13%
h. Top PI	14%	16%	12%	17%	12%	0%	7%	14%	16%	22%(2)
i. Other	1%	0%	2%	0%	2%					
Total	263	83	180	77	186	6	55	114	56	32

*As reported by the researchers.
**First five or less papers published within contexts (a) to (h).
(X) Relative ranking within category.

Table 4.8 provides the response rate for the open question "what opportunities did these different environments open?" and followed the same structure of the last table. From this table it can be seen that access to economic resources is the most important avenue that the different research environments have opened to young scientists. In addition, when comparing star and non-star researchers, there seems to be some difference between the two. The former ranks exposure to other ideas in second place and gives more importance to collaboration with foreigners, whereas the latter ranks exposure to other ideas in third place. Additionally, SNI top tier gave a greater importance to exposure to other ideas than any other option. In contrast, less prominent scientists and the lower ranks of SNI gave more importance to access to specialized laboratories. These outcomes suggest that non-star researchers focus more on resources, whereas eminent ones are more about ideas and international exposure.

Table 4.8. Perceived Opportunities* of Research Environments upon entry and Eminence

Opportunities (percentage within category)		Researcher sees herself as a star								
		Based on Art.		Based on Cit.		SNI Level				
	All	Yes	No	Yes	No	None	Candidate	Level 1	Level 2	Level 3
Access to										
i. Economic resources	36%(1)	30%(1)	38%(1)	32%(1)	37%(1)	50%	45%(2)	35%(1)	32%(2)	25%(3)
ii. Qualified personnel	17%	17%	17%	18%	16%	17%	20%	17%	13%	19%
iii. Specialized laboratories	28%(2)	18%	32%(2)	19%	31%(2)	0%	47%(1)	25%(3)	21%	19%
Collaboration	0%	0%	1%	0%	1%	0%	0%	0%	2%	0%
ii. w/ labs	16%	17%	15%	18%	15%	17%	7%	17%	21%	16%
iii. w/ universities	18%	18%	18%	19%	18%	17%	16%	20%	18%	16%
iv. w/ foreigners	25%	25%(3)	24%	27%(3)	24%	17%	16%	24%	36%(1)	28%(2)
g. Exposed to other ideas	27%(3)	27%(2)	28%(3)	29%(2)	27%(3)	17%	22%(3)	31%(2)	25%(3)	31%(1)
Total	263	83	180	77	186	6	55	114	56	32

*As reported by the researchers
(X) Relative ranking within category.

Table 4.9 provides the response rate for the question “what have you learned from the different environments (stated previously)?[51],” where we follow the same structure of the last two tables. In general, there seems to be a consensus among the different categories that two of the main things scientists have leaned from the different settings is to (1) collaborate with other authors and (2) have access to new research techniques and (3) publish in ISI journals. However. Similarly to what was seen above, there seems to be a significant difference between researchers that have average productivity (or impact) vis-à-vis their eminent peers. Star scientists (as well as SNI levels 2 and 3) think that exposure to new ideas is the main insight they have learned from the different research environments, whereas non-star scientists give a higher weight to exposure to new research techniques. In addition, the data suggest that the higher the SNI level the higher

[51] For this question we received different types of answers, ranging from one or two words to small paragraphs describing what scientists learned within one or more researcher settings. All the answers were analyzed and converted to fit one of the following options: a) new ideas, b) new research techniques, c) write research proposals, d) publish ISI articles, e) teach/advise scientists, and d) collaborate with other scientists. In some cases, there was a 1-to-1 relationship between the answer and a setting but in others an answer could fit more than one environment, in which case we included all the options that were relevant.

the importance of learning new ideas in their early collaborations; ranging from 25% in candidates up to 56% in SNI's top tier. Furthermore, these results show that there is a difference between scientists that see themselves as highly productive and the ones that have high impact; the former learned how to publish ISI articles in their initial collaborations, whereas the latter learned new research techniques. These results suggest that less preeminent scholars concentrate more on acquiring the "basic" research skills (like learning how to publish or understanding the different research techniques) from their early collaborators. In contrast, top scientists center their efforts on exposure to new ideas and the frontier of science.

Table 4.9. Learning* within a Research Environment upon entry and Eminence

Learning (percentage within category)	All	Researcher sees herself as a star				SNI Level				
		Based on Art.		Based on Cit.						
		Yes	No	Yes	No	None	Candidate	Level 1	Level 2	Level 3
a. New ideas	38%	41%(1)	35%	48%(1)	31%	33%	25%	37%	43%(1)	56%(1)
b. New research techniques	41%(1)	33%	44%(1)	38%(3)	40%(1)	33%	45%(2)	43%(2)	36%(2)	41%(2)
c. Write research proposals	24%	18%	26%	22%	22%	50%	25%	28%	14%	19%
d. Publish ISI articles	39%(3)	36%(2)	39%(2)	35%	37%(2)	17%	49%(1)	42%(3)	29%	31%(3)
e. Teach/advise scientists	15%	16%	15%	17%	15%	17%	20%	18%	9%	9%
d. Collaborate w/ other scientists	40%(2)	41%(1)	39%(2)	47%(2)	35%(3)	33%	38%(3)	45%(1)	43%(1)	25%
Total	263	83	180	77	186	6	55	114	56	32

*As reported by the researchers
(X) Relative ranking within category.

4.5. Conclusion and Policy Implications

In the last decades, ST&I have been seen as a major source of economic growth. A key and growing force behind the evolution of science is the ability of scientists to collaborate with other researchers, foster nurturing environments and develop cooperative projects with other sectors.

In the past, scholars have studied the factors that promote research collaboration and assessed the cost and benefits these interactions produce for the individual researcher and the whole realm of science. This paper contributes to this body of knowledge by evaluating how research collaboration in the early stages of the academic life of a researcher can influence the development of new researchers. In particular, this work assesses the impact that various research environments have on the development of the next generation of scientists, as well as the factors that promote collaboration with a particular setting. In addition, it looks into the opportunities these interactions open and benefits they produce to this group of people.

Our analysis confirms the importance of research collaboration in the development of a scientific system, especially within the context of an emerging economy like Mexico. In particular, this work shows that 95% of physicists in Mexico[52] started early on in their career cooperating with another scientist, and three quarters did so in the context of a research group. In addition, this work shows that scientists highly regard their initial collaborators, since 47% to 73% said that early on in their careers they conducted research within a top research environment, like an eminent scientist or top research group, during their first publications.

This works also shows that scientists will have a higher propensity to start collaborating organically with an informal group of researchers, in this case two to six co-authors. But if they have to apply for a position, like a Ph.D. or postdoctoral research program, their initial interactions will be with an established RG. Additionally, proactive scholars (i.e., the ones that actively sought out their co-authors) will prefer to collaborate with the leaders of those groups. These observations suggest that competitive processes and the entry of new scientists in the

[52] Our sample consisted in 263 researchers; this represents almost 20% of all the SNI researchers that belong to this field and related areas.

context of formal research groups are two related and intertwined dimensions in the evolution of a science system.

In terms of the impact research collaboration has on new scientists, this work confirms the importance that eminent researchers, as well as top PIs, have in the development of other prominent scholars, in particular in SNI's top tier. In addition, this study suggests that star and non-star scientists, early on in their academic careers, focus on opening different opportunities and gaining different knowledge from their interactions with other researchers. In particular, less preeminent researchers focus more on gaining access to economic and physical resources (like specialized laboratories), and learning "basic" scientific skills (like publishing or research techniques), whereas more preeminent ones are more about new ideas and international exposure.

These results have implications for policy-making in science, technology and innovation systems. They deepen our understanding of the role that eminent scientists, but also strong research groups, have in developing the next generation of prominent scholars. In addition, they suggest that initial collaborations in science open different opportunities and teach different things to star and non-star scientists. In particular, early interactions allow prominent scholars to gain access to international scholars and learn the latest ideas. This means that if a system wants to be on the cutting edge of science it should actively advance international collaboration, create exchange agreements for their young scientists, and promote the acquisition of nascent ideas. While the traditional way has been through the apprenticeship at the lab, it may be possible to provide access to these new and important ideas through alternative mechanisms.

Chapter 5. Conclusions

As research collaboration increases in importance, research administrators and policy makers are trying to better grasp this phenomenon. Changes in the realm of science, including tightening budgets and higher awareness of the outputs of ST&I systems, are demanding a better understanding of cooperation and improved methods for assessing these interactions. This thesis exploits patterns of collaboration in science to propose and test a method that identifies and assesses research groups endogenously. In addition, it evaluates the impact this cooperation has on the progress of the ST&I system. In particular, it considers the influence that collaborative environments, including research groups or the presence of a prominent scientist, have on new scholars, as well as the opportunities that these various milieus open to the acquisition of new knowledge and competencies. This concluding chapter summarizes the major findings and discusses its contributions to the literature, its implications for science policy, and its implications for future research.

5.1. Summary of empirical findings

Using the patterns of collaboration and the strength of these interactions, Chapter 2 proposes a method to identify and assess research groups that takes into account the self-organizing characteristics of the research endeavor. The method is then tested with a database from the fields of Physics and related areas containing all the papers published in Mexico between 1995 and 1999 (as reported by ISI). The method produces three main results. First, the strength and frequency of the collaboration patterns allow us to identify cohesive groups, regardless of the institutional or location context of their members (researchers). In addition, this new technique allows its users to take into account the lack of homogeneity within institutions in their evaluation efforts. Second, the knowledge footprint (KFP) allows potential evaluators to identify similar research groups, assess these groups and produce more meaningful evaluations, where benchmarks have reasonable proximity in their research characteristics. Third, the method shows that research done by the different groups in Physics and related areas in Mexico is mostly non-redundant. This means that each RG is (virtually) focused in one area of the research space with little overlap with other groups. This is perhaps not surprising because of the small size of the Mexican science system, which provides ample opportunities for each group to find its own research space.

Chapter 3 provides quantitative evidence that prominent scholars have a prime role in the development of the scientific system, especially within the context of an emerging economy like Mexico. More significantly, this group of scientists has a significant effect on the research productivity and impact of new researchers with whom they collaborate upon entry into science. Their contribution is quite significant, with this early link conditioning the likelihood that young faculty will in turn become a future star. This chapter also shows that nurturing environments, like top research groups, have an impact on the scientific outcomes of incoming scientists, but only in terms of their productivity (i.e. number of publications), not necessarily on their impact and visibility (as measured by citations).

Finally, Chapter 4 builds on the previous section by asking researchers to assess how the various nurturing environments, including prominent scholars, top research groups and other types of milieus impact the opportunities and acquired knowledge of young researchers. This chapter suggests that, from the beginning of their career, star scientists have a different mindset when compared with non-star scientists. Prominent scholars are focused on acquiring new ideas and being exposed to the frontier of science through international collaboration and relations with existing stars, whereas less prominent scholars are more focused on obtaining access to economic and physical resources (like specialized laboratories), and learning “basic” scientific skills (like publishing or research techniques).

5.2. Policy Implications

A variety of science and public policy challenges arise from the findings of the thesis. First, when making assessments of the ST&I, this thesis suggests that research administrators, policy makers and scholars should be aware of the heterogeneity within this system. This means that they should pay close attention to how they define the unit of analysis, and compare and rank these entities only when there is a reasonable overlap in their knowledge footprint. This is particularly important in small emerging science systems, as in the case of Mexico, where research groups are likely to be very heterogeneous, with productive and lagging groups working side by side in the same department. In such cases, an average level of scientific productivity for an overall departmental unit produces information of limited value to the assessors. Assessments should incorporate international visiting committees, and the performance of the different units of analysis may be compared with peers in similar economies/systems, rather than unrelated units within the same system. For example, for the Mexican research groups, a benchmark could be Brazilian teams.

Second, the findings of the thesis are consistent with existing research suggesting that prominent scientists have a primary role in the progress of the ST&I system. But the work goes further. First it shows the critical role that these stars play in the development and productivity of young faculty. In addition, it demonstrates how research collaborators in the context of a group play an additional role in the construction of these systems, especially to learn the tools of the trade. This means that research groups are key for the dissemination of the scientific endeavor and cumulatively building of domestic capabilities. But if a country or region wants to become a leader in a certain area of knowledge, it may need to focus on attracting and retaining the best and the brightest, and creating around these key figures appropriate collaborative environments so that top research can flourish.

Finally, the results of this thesis show that, early on in their career, scholars who later become stars benefit by collaborating with international peers and existing star scientists, which allows them to be exposed to the frontier of science by learning new ideas. In contrast, less prominent scientists focus more on obtaining access to economic and physical resources (like specialized laboratories), and learning "basic" scientific skills (like publishing or research techniques) through their research groups. Therefore, if a system wants to promote the development of its researchers, particularly eminent ones, international collaboration is critical, and the exchange of scientists across and within countries and institutions, (especially with leading research centers), appears to be particularly significant for young scholars. What they get through this exposure are the new ideas of the future of science. That, rather than just the how to or the resources, appears to distinguish the good from the top.

5.3. Future work

While the analyses and results of this thesis fill important gaps on the assessment of science and research collaboration literature, they also generate new questions that need to be addressed in subsequent work. First, further work should look at the effect that other measures of group cohesiveness (e.g. n-cliques, k-plexes, etc.) have on collaborative groups. In addition, additional analysis is needed to test the robustness of the method shown in chapter 2, by incorporating other fields of knowledge or using data from other countries or regions. Furthermore, this method could be used to increase our understanding of the determinants of research productivity (Gonzalez-Brambila and Veloso, 2007) in a number of ways. One possibility is to study how the characteristics of the naturally emerging groups are tied to their productivity. Another possibility would be to extend the approach to other types of research output data amenable to equivalent analysis, in particular patents.

Second, further analysis on the impact eminent scientists have on the ST&I system should incorporate a variety of complementary variables. One would certainly be the availability and accessibility of individual and institutional economic resources. Another is the role that the star plays within the research organization, distinguishing whether this person is a department head or the leader of a lab, as well as their position within its broader network of collaborators and the whole system (e.g. degree centrality, structural hole). Country of origin and, in general, international vs. domestic collaborations are also a dimension that could be further brought to complement the work. Finally, while ISI papers and citations have been the core variables used in the research, it would be good to include other measures of performance, such as number of Master and Ph.D. advisees, or patents. Each and all of these would allow a more complete, and therefore more robust, picture of the role of eminences and research groups.

References

Adams, J., 1998. Benchmarking International Research. Nature 396, 615-618.

Adams, J.D., Black, G.C., Clemmons, J.R., Stephan P.E., 2005. Scientific teams and institutional collaborations: Evidence from U.S. universities, 1981–1999. Research Policy 34, 259-285.

Allison, P.S., Stewart, J.A., 1974. Productivity Differences Among Scientists: Evidence for Accumulative Advantage. American Sociological Review 39, 596-606.

Allison, P.D., Long, J.S., 1990. Departmental Effects on Scientific Productivity. American Sociological Review 55, 469-478.

Allen, T.D., Eby, L.T., Poteet, M.L., Lentz, E., Lima, L., 2004. Career benefits associated with mentoring for protégé: A meta-analysis. Journal of Applied Psychology 89, 127-136.

Amabile, T.M., Gryskiewicz, N.D., 1989. The creative environment scales: The work environment inventory. Creativity Research Journal 2, 231–254.

Amabile, T.M., Patterson, C., Mueller, J., Wojcik, T., Odomirok, P.W., Marsh, M., Kramer, S.J., 2001. Academic-practitioner collaboration in management research: A case of cross-profession collaboration. The Academy of Management Journal, 418–431.

Azoulay P., Zivin J.G., Wang J., 2010. Superstar Extinction. The Quarterly Journal of Economics 125, 549-589.

Australian Research Council, 2001. Federation Fellowships. Discovery Newsletter, 1–2.

Bammer, G., 2008. Enhancing research collaborations: Three key management challenges. Research Policy 37, 875–887.

Barabasi, A.L., Jeong, H., Neda1, Z., Ravasz, E., Schubert, A., Vicsek, T., 2002. Evolution of the social network of scientific collaborations. Physica A: Statistical Mechanics and its Applications 311, 590-614.

Bargar, R.R., Mayo-Chamberlain, J., 1983. Advisor and advisee issues in doctoral education. The Journal of Higher Education 54, 407-432.

Bertrand, M., Mullainathan, S., 2001. Do people mean what they say? Implications for subjective survey data. American Economic Review 91, 67–72.

Birnholtz, J.P., 2007. When do researchers collaborate? toward a model of collaboration propensity. Journal of the American Society for Information Science and Technology 58, 2226–2239.

Blackburn, R.T., Chapman, D.W., Cameron, S.M., 1981.'Cloning' in academe: Mentorship and academic careers. Research in Higher Education 15, 315-327.

Bland, C.J., Ruffin, M.T., 1992. Characteristics of a productive research environment- literature review, Academic Medicine 67, 385-397.

Braun, T., Gómez, I., Méndez, A., Schubert, A., 1992. International co-authorship patterns in physics and its subfields, 1981–1985. Scientometrics 24, 181-200.

Bozeman, B., Dietz, J., Gaughan, M., 2001. Scientific and technical human capital: an alternative model for research evaluation. International Journal of Technology Management 22, 636– 655.

Bozeman, B., Corley. E., 2004. Scientists' collaboration strategies: implications for scientific and technical human capital. Research Policy 33, 599-616.

Braxton, J.M., 1983. Department Colleagues and Individual Faculty Publication Productivity. Review of Higher Education 6, 115-128.

Bukvova, H., 2010. Studying Research Collaboration: A Literature Review. Sprouts: Working Papers on Information Systems 10, http://sprouts.aisnet.org/10-3.

Cable, D.M., Murray, B., Tournaments versus Sponsored Mobility as Determinants of Job Search Success. The Academy of Management Journal 42, 439-449.

Callon, M., Courtial, J.P., Turner, W.A., Bauin, S., 1983. From translations to problematic networks: an introduction to co-word analysis. Social Science Information 22, 191–235.

Cameron, S.W., Blackburn, R.T., 1981. Sponsorship and Academic Career Success. The Journal of Higher Education 52, 369-377.

Carley, K.M., 2002. Smart Agents and Organizations of the Future, in: Lievrouw, L., Livingstone, S., (Eds.), The Handbook of New Media: Social Shaping and Consequences of Icts, Sage, Thousand Oaks, CA, pp. 206-220.

Carley, K.M., 2003. Dynamic Network Analysis, in: Breiger, R., Carley, K.M., Pattison, P. (Eds), Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers. National Academies Press, Washington D.C., pp. 133-145.

Cervantes, M., 2004. Attracting, retaining and mobilising high skilled labour, in OECD (Ed.), Global Knowledge Flows and Economic Development, OECD, Paris, pp. 51-71.

CINVESTAV, 2008. Jerzy Plebanski's Biography (downloaded on 12 January 2009 from http://www.fis.cinvestav.mx/~grupogr/gr_mphysics/plebanski/Plebanski.html).

Cole, J.R., 1970. Patterns of intellectual influence in scientific research. Sociology of Education 43, 377-403.

Cole, J.R., Cole, S., 1973. Social Stratification in Science. University of Chicago Press, Chicago, IL.

Cole, J.R., Cole, S., 1972. The Ortega Hypothesis. Science 178, 368-375.

Cole, J.R., 1979. Fair Science: Women in the Scientific Community. The Free Press, New York.

Collazo-Reyes, F., Luna-Morales, M.E., 2002. Física mexicana de partículas elementales. INCI 27, 347-353.

Collazo-Reyes, F., Luna-Morales, M.E., Russell, J.M., 2004. Publication and citation patterns of the Mexican contribution to a “Big Science” discipline: Elementary particle physics. Scientometrics 60, 131-143.

CONACYT, 2004. Informe General del Estado de la Ciencia y Tecnología. Consejo Nacional de Ciencia y Tecnología, Mexico City.

CONACYT, 2008. Science and Technology Indicators: At Glance. Consejo Nacional de Ciencia y Tecnología, Mexico City.

CONACYT, 2011. Sistema Nacional de Investigadores, Investigadores viegentes 2011. (downloaded on 15 March 2011 from http://www.conacyt.mx/SNI/Documents/SIN-investigadores-vigentes-2011.pdf).

CONACYT-SIICYT, 2011. Indicadores Cientificos y Tecnologicos, Sistema Integrado de Informacion Cientifica y Tecnologica (downloaded on 15 November 2011 from www.siicyt.gob.mx).

CONACYT, 2012. Sistema Nacional de Investigadores – SIN. (downloaded on 12 March 2012 from http://www.conacyt.gob.mx/SNI/Paginas/default.aspx).

COSEPUP, 1999. Evaluating Federal Research Programs: Research and the Government Performance and Results Act. National Academy Press, Washington, DC.

Coulter, N., Monarch, I., Konda, S., 1998. Software Engineering as Seen through Its Research Literature: A Study in Co-Word Analysis. Journal of the American Society for Information Science 49,1206–1223.

Crane, D., 1965. Scientists at Major and Minor Universities: A Study of Productivity and Recognition. American Sociological Review 30, 699-714.

Crane. D., 1972. The Invisible College. University of Chicago Press, Chicago, IL.

Cronan-Hillix, T., Gensheimer, L.K., Cronan-Hillix W.A., Davidson, W.S., 1986. Students' views of mentors in psychology graduate training. Teaching of Psychology 13, 123-127.

Cummings, J., Kiesler, S., 2007. Coordination costs and project outcomes in multi-university collaborations. Research Policy 36, 1620-1634.

de Solla Price, D.J., 1963. Little Science, Big Science. Columbia University Press, New York.

de Solla Price, D.J., Beaver, D.B., 1966. Collaboration in an invisible college. American Psychologist 21, 1011–1018.

deB. Beaver, D., Rosen, R., 1978. Studies in scientific collaboration Part I. The professional origins of scientificco-authorship. Scientometrics 1, 65-85.

deB. Beaver, D., Rosen, R., 1979a. Studies in scientific collaboration Part II. Scientific co-authorship, research productivity and visibility in the French scientific elite, 1799–1830. Scientometrics 1, 133-149.

deB. Beaver, D., Rosen, R., 1979b. Studies in scientific collaboration Part III. Professionalization and the natural history of modern scientific co-authorship. Scientometrics 1, 231-245.

deB. Beaver, D., 2001. Reflections on scientific collaboration (and its study): past, present, and future. Scientometrics 53, 365–377.

Defazio, D., Lockett, A., Wright, M., 2009. Funding incentives, collaborative dynamics and scientific productivity: Evidence from the EU framework program. Research Policy 38, 293–305.

Dias, M.B., Reyes-Gonzalez, L., Veloso, F.M., Casman, E.A., 2010. Effects of the USA PATRIOT Act and the 2002 Bioterrorism Preparedness Act on select agent research in the United States. Proceedings of the National Academy of Sciences 07, 9556-9561.

Dill, D.D., 1985. Theory versus practice in the staffing of R&D laboratories. R&D Management 15, 227-241.

Dill, D.D., 1986. Research as a scholarly activity: Context and culture. New Directions of Institutional Research 50, pp. 7-23.

Ding, W., Choi, E., 2011. Divergent paths to commercial science: A comparison of scientists' founding and advising activities. Research Policy 40, 69-80.

Donaldson, S.I., Ensher, E.A., Grant-Vallone, E.J., 2000. Longitudinal examination of mentoring relationships on organizational commitment and citizenship behavior. Journal of Career Development 26, 233-249.

Duque, R.B., Ynalvez, M., Sooryamoorthy, R., Mbatia, P., Dzorgbo, D.B.S., Shrum, W., 2005. Collaboration Paradox: Scientific Productivity, the Internet, and Problems of Research in Developing Areas. Social Studies of Science 35, 755-785.

Etzkowitz, H., Leydesdorff, L., 2000. The dynamics of innovation: from National Systems and "Mode 2" to a Triple Helix of university–industry–government relations. Research Policy 29, 109-123.

Etzkowitz, H., 2001. The Second Academic Revolution and the Rise of Entrepreneurial Science. IEEE Technology and Society 22, 18-29.

Fagenson, E.A., 2006. The mentor advantage: Perceived career/job experiences of protégés versus non-protégés. Journal of Organizational Behavior 10, 309-320.

Ford, J.R., Fulkerson, D.R., 1962. Flows in Networks. Princeton University. Press, Princeton, NJ.

Frederiksen, L.F., Hansson, F., Wenneberg, S.B., 2003. The Agora and the Role of Research. Evaluation 9, 149–172.

Fox, M.F., 1983. Publication Productivity among Scientists: A Critical Review. Social Studies of Science 13, 285-305.

Fox, M.F., 2001. Women, Science, and Academia: Graduate Education and Careers. Gender and Society 15, 654-666.

Fox, M.F., Mohapatra, S., 2007. Social-Organizational Characteristics of Work and Publication Productivity among Academic Scientists in Doctoral-Granting Departments. The Journal of Higher Education 78, 542-571.

Garfield, E., 1979. Is Citation Analysis a Legitimate Evaluation Tool? Scientometrics 1, 359-375.

Georghiou, L., Roessner, D., 2000. Evaluating technology programs: tools and methods. Research Policy 29, 657–678.

Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P., Trow, M., 1994. The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies. Sage Publications Ltd, London, UK.

Girvan, M., Newman, M.E.J., 2002. Community structure in social and biological networks. Proceedings of the National Academy of Sciences 99, 7821–7826.

Glänzel, W., Schubert, A., 2004. Analyzing Scientific Networks through co-authorship, in: Glänzel, W., Schmoch U. (Eds.), Handbook of Quantitative Science and Technology Research, Springer, Dordrecht, The Netherlands, pp. 257-276

Gläser, J., Spurling T.H., Butler, L., 2004. Intraorganisational evaluation: are there ‘least evaluable units’? Research Evaluation 13, 19-32.

Glasow, P.A., 2005. Fundamentals of Survey Research Methodology. MITRE Case #05-0638 (downloaded on March 5, 2012 from www.mitre.org/work/tech_papers/tech_papers_05/.../05_0638.pdf).

Gmür, M., 2003. Co-citation analysis and the search for invisible colleges: A methodological evaluation. Scientometrics 57, 27 – 57.

Gomez, I., Fernandez, M.T., Sebastian, J., 1999. Analysis of the structure of international scientific cooperation networks through bibliometric indicators. Scientometrics 44, 441-457.

Gonzalez-Brambila, C.N., and Veloso, F.M., 2007. The determinants of research output and impact: A study of Mexican researchers. Research Policy 7, 1035–1051.

Goodall, A.H., 2009., Highly cited leaders and the performance of research Universities. Research Policy 38, 1079-1092.

Green, S.G., 1991., Professional Entry and the Adviser Relationship: Socialization, Commitment, and Productivity. Group & Organization Management 16, 387-407.

Green, S.G., Bauer, T.N., 1995. Supervisory Mentoring by Advisers: Relationship with Doctoral Students Potential, Productivity and Commitment. Personnel Psychology 48, 537-562.

Green, W.H., 2002. Econometric analysis, 5th ed, Prentice-Hall, New Jersey, NJ.

Guston, D.H., 2003. The expanding role of peer review process in the United States, in: Shapira, P., Kuhlmann, S. (Eds.), Learning from Science and Technology Policy Evaluation. Edward Elgar Publishing, Northampton, MA, pp. 81-97

Guimera, R., Uzzi, B., Spiro, J., Nunes Amaral, L.A. 2005. Team Assembly Mechanisms Determine Collaboration Network Structure and Team Performance. Science 308, 697-702.

Ham, J.C., Weinberg, B.A., 2007. Geography and Innovation: Evidence from Nobel Laureates. Working Paper, Ohio State University.

Hagstrom, W.O., 1968. Departmental prestige and scientific productivity. Paper presented at the 63rd Annual Meeting of the American Sociological Association, Boston, Mass.

Hausman, J., Taylor, T., 1981. Panel Data and Unobservable Individual Effects. Econometrica 49, 1377-98.

Horta, H., Veloso, F.M., Grediaga, R., 2010. Navel Gazing: Academic Inbreeding and Scientific Productivity. Management Science 56, 414-429.

He, Z., Geng. X., Campbell-Hunt, C., 2009. Research collaboration and research output: A longitudinal study of 65 biomedical scientists in a New Zealand university. Research Policy 38, 306-317.

Ham, JC, and Weinberg, BA. (2007). "Geography and Innovation: Evidence from Nobel Laureates." Working Paper, Ohio State University.

Hemlin, S., Allwood, C.M., Martin, B.R., 2009. Creative Knowledge Environments. Creativity Research Journal 20, 196–210.

Hummon, N.P., Carley, K., 1993. Social networks as a normal Science. Social Networks 15, 71-106.

Hunter, S.T., Bedell, K.E., Mumforda, M.D., 2007. Climate for Creativity: A Quantitative Review. Creativity Research Journal 19, 69-90

Institute of Scientific Information, 2003. National Citation Report for Mexico. Articles from ISI database years 1981–002 cited in the period 1981–002, Project 2487, April 18.

Jain, S., George, G., Maltarich, M., 2009. Academics or entrepreneurs? Investigating role identity modification of university scientists involved in commercialization activity Research Policy 38, 922-935.

Jeffrey, P., 2003. Smoothing the waters: Observations on the process of cross-disciplinary research collaboration. Social Studies of Science 33, 539– 562.

Judge, T.A., Kammeyer-Mueller, J., Bretz, R.D., 2004. A longitudinal model of sponsorship and career success: A study of industrial-organizational psychologists. Personnel Psychology 57, 271-303.

Kane, A., 2001. Indicators and Evaluation for Science, Technology and Innovation. ICSTI Task Force on Metrics and Impact (downloaded on 15 August 2005 from http://www.economics.nuigalway.ie/people/kane/2002_2003ec378.html).

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

Kennedy, D., 2003. Multiple Authors, Multiple Problems. Science 301, 733.

Kernighan, B.W., Lin, S., 1970. An efficient heuristic procedure for partitioning graphs. Bell System Technical Journal 49, 291–307.

King, A.D., 2004. The Scientific Impact of Nations. Nature 430, 311-316.

Krackhardt, D., Carley, K.M., 1998. A PCANS Model of Structure in Organization. Proceedings of the 1998 International Symposium on Command and Control Research and Technology. Monterray, CA. Evidence Based Research, pp. 113-119.

Krackhardt, D., 1999. The ties that torture: Simmelian tie analysis in organizations. Research in the Sociology of Organizations, 183–210.

Kram, K.E., Glenview, I.L., Foresman, S., 1985. Mentoring at work: Developmental relationships in organizational life. Human Resource Management 25, 639–644.

Kuhn, T.S., 1962. The Structure of Scientific Revolutions. University of Chicago Press, Chicago, Il.

Laudel, G., 1999. Interdisziplinäre Forschungskooperation: Erfolgsbedingungen der Institution 'Sonderforschungsbereich'. Edition Sigma, Berlin.

Laudel, G., 2001. Collaboration, creativity and rewards: why and how scientists collaborate. International Journal of Technology Management 22, 762-781.

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

Laudel, G., 2005. Migration Currents Among the Scientific Elite. Minerva 43, 377–395.

Lee, L., Bozeman, B., 2005. The Impact of Research Collaboration on Scientific Productivity. Social Studies of Science 35, 673-702.

Leydesdorff, L., 1987. Various methods for the Mapping of Science. Scientometrics 11, 291-320.

Leydesdorff, L., Etzkowitz, H., 2001. The Transformation Of University-industry-government Relations. Electronic Journal of Sociology 5 (downloaded on January 30, 2012 from http://www.sociology.org/content/vol005.004/th.html).

Levin, G.L., Stephan, P.E., 1998. Gender Differences in the Rewards to Publishing in Academe: Science in the 1970s. Sex Roles 38, 1049-1064.

Lissenburgh, S., Rebecca, H., 2000. Knowledge Links: Innovation in university/business partnerships. IPPR, London.

Long, J.S., 1978. Productivity and Academic Position in the Scientific Career. American Sociological Review 43, 889-908.

Long, J.S., Allison, P.D., McGinnis, R., 1979. Entrance into the Academic Career. American Sociological Review 44, 816-830.

Long, J.S., McGinnis, R., 1981. Organizational contexts and scientific productivity. American Sociological Review 46, 422-442.

Long, J.S., Fox, M.F., 1995. Scientific Careers: Universalism and Particularism. Annual Review of Sociology 21, 45-71.

Louis, K.S., Holdsworth, J.M., Anderson, M.S., Campbell, E.G., 2007. Becoming a Scientist: Correction The Effects of Work-Group Size and Organizational Climate. The Journal of Higher Education 78, 311-336.

Lotka A., 1926. The frequency distribution of scientific productivity. Journal of the Washington Academy of Sciences 16, 317–323.

Louis, K.S., Holdsworth, J.M., Anderson, M.S., Campbell, E.G., 2007. Becoming a Scientist: Correction The Effects of Work-Group Size and Organizational Climate. The Journal of Higher Education 78, 311-336.

Luukkonena, T., 2001. Networking Impacts of the EU Framework Programme. Innovative networks: co-operation in national innovation systems, OECD, Paris.

Malmgren, R.D., Ottino, J.M., Nunes Amaral, L.A., 2010. The role of mentorship in protégé performance. Nature, 465, 622-626.

Martin, B.R., Irvine, J., 1983. Assessing basic research: Some partial indicators of scientific progress in radio astronomy. Research Policy 12, 61-90.

Marx, W., Hoffmann, T., 2011. Bibliometric analysis of fifty years of physica status solidi. Physica Status Solidi, Forthcoming.

May, R.M., 1997. The Scientific Wealth of Nations. Science 275, 793–796.

McIntyre, L.J., 1999. The practical skeptic: Core concepts in sociology. Mayfield Publishing, Mountain View, CA.

Melin, G., Persson, O., 1996. Studying research collaboration using co-authorships. Scientometrics, 36, 363-377.

Melin, G., 2000. Pragmatism and Self-organization: Research Collaboration on the Individual Level. Research Policy 29, 31–40.

Merton, R.K., 1973. The Sociology of Science. The University of Chicago Press, Chicago, IL.

Miquel, J.F., Ojasoo, T., Okubo, Y., Paul, A., J. Doré, C., 1995. World science in 18 disciplinary areas: Comparative evaluation of the publication patterns of 48 countries over the period 1981–1992. Scientometrics 33, 149-167.

Mundlak, Y., 1978. On the Pooling of Time Series and Cross Section Data. Econometrica 46, 69-85.

Moed, H.F., 2005. Citation Analysis in Research Evaluation. Springer, Dordrecht, The Netherlands.

Nagpaul, P.S., Sharma, L., 1994. Research output and transnational cooperation in physics subfields: A multidimensional analysis. Scientometrics 31, 97-122.

Narin, F., 1976. Evaluative Bibliometrics: The Use of Publication and Citation Analysis in the Evaluation of Scientific Activity. National Science Foundation, Washington D.C.

Narin, F., 1978. Objectivity versus relevance in studies of scientific advance, Scientometrics 1, 35-41.

Narvaez-Berthelemot, N., Frigoletto, L.P., Miquel, J. F., 1992. International scientific collaboration in Latin America. Scientometrics 24, 373-392.

Nelson, R., 1993. National Innovation Systems: A Comparative Analysis. Oxford University Press, Oxford.

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., 2003. Fast algorithm for detecting community structure in networks. Physical Review E 69, 066133.

Newman, M.E.J., 2004a. Detecting community structure in networks. The European Physical Journal B - Condensed Matter and Complex Systems, 38, 321-330.

Newman, M.E.J., 2004b. Coauthorship networks and patterns of scientific collaboration. Proceeding of the National Academy of Science 101, 5200–5205.

Newman, M.E.J., 2006. Modularity and community structure in networks. Proceeding of the National Academy of Science 103, 8577–8582.

Newman, M.E.J., Girvan, M., 2003. Finding and evaluating community structure in networks. Physical Review E 69, 026113.

Noyons, E.C.M., Moed, H.F., Luwel, M., 1999. Combining Mapping and Citation Analysis for Evaluative Bibliometric Purposes: A Bibliometric Study. Journal of the American Society for Information Science 50, 115–131.

Noyons, E.C.M., 2004. Science maps within a science policy context, in: Moed, H.F., Glänzel, W., Schmoch, U. (Eds), Handbook of quantitative science and technology research. The use of publication and patent statistics in studies of S&T systems. Kluwer Academic Publishers, Dordrecht, pp. 237–256.

OECD, 1999. Managing national innovation systems. OECD, Paris.

OECD, 2001. Innovative Networks Co-operation in National Innovation Systems. OECD, Paris.

OECD, 2004a. Global Knowledge Flows and Economic Development. OECD, Paris.

OECD, 2004b. OECD Science, Technology and Industry Outlook 2004. OECD, Paris.

OECD, 2011. OECD Science, Technology and Industry Outlook 2010. OECD, Paris.

Oettl A., 2009. Productivity and Helpfulness: A New Taxonomy for Star Scientists. Working Paper.

Ordoñez-Matamoros, G., Cozzens, S.E., García-Luque, M., 2009. Who Collaborates Internationally in Developing Countries?: The Case of Colombia. paper presented at the 2009 Gobelics Conference in Dakar, Senegal.

Paglis, L.L., Green, S.G., Bauer, T.N., 2006. Does Adviser Mentoring Add Value? A Longitudinal Study of Mentoring and Doctoral Student Outcomes. Research in Higher Education 47, 451-476.

Palla, G., Derényi, I., Farkas, I, Vicsek, T., 2005. Uncovering the overlapping community structure of complex networks in nature and society. Nature 435, 814-818.

Papaconstantinou, G., Polt, W., 1997. Policy Evaluation and Technology: An Overview. OECD Proceedings, Conference on Policy Evaluation in Innovation and Technology. OECD, Paris.

Payne, S.C., Huffman, A.H., 2005. A longitudinal examination of the influence of mentoring on organizational commitment and turnover. Academy of Management Journal 48, 158–168.

Persson, O., Glänzel, W., Danell, R., 2004. Inflationary bibliometric values: The role of scientific collaboration and the need for relative indicators in evaluative studies. Scientometrics 60, 421-432.

Phillips, G.M., 1979. The peculiar intimacy of graduate study: A conservative view. Communication Education 28, 339-345.

Pickrell, J., 2001. Fighting Brain Drain: Ireland Gives Its Stars a Big Pot o' Gold. Science 293, 1028-1029.

Pinsonneault, A., Kraemer, K.L., 1993. Survey research methodology in management information systems: An assessment. Journal of Management Information Systems, 10, 75-105.

Ramsdena, P., 1998. Managing the Effective University. Higher Education Research & Development 17, 347-370.

Radicci, F., Castellano, C., Cecconi, F., Loreto, V., Parisi, D., 2004. Defining and identifying communities in networks. Proceeding of the National Academy of Science 101, 2658–2663.

Radicchi, F., Castellano, C., Cecconi, F., Loreto, V., Parisi, D., 2004. Defining and identifying communities in networks. Proceedings of the National Academy of Sciences 101, 2658–2663.

Ravasz, E., Somera, A.L., Mongru, D. A., Oltvai, Z., Barabasi, A.L., 2002. Hierarchical organization of modularity in metabolic networks. Science 297, 1551–1555.

Reskin, B.F., 1977. Scientific Productivity and the Reward Structure of Science. American Sociological Review 42, 491-504.

Reskin, B.F., 1978. Scientific Productivity, Sex, and Location in the Institution of Science. American Sociological Review 83, 1235-43.

Reskin, B.F., 1979. Academic Sponsorship and Scientists' Careers. Sociology of Education 52, 129-46.

Rip. A., 2000. Societal Challenges for R&D Evaluation, in Shapira, P., Kuhlmann, S. (Eds), Learning from Science and Technology Policy Evaluation. Edward Elgar Publishing, Cheltenham, UK, pp. 18-41.

Rigby, J., Edler, J., 2005. Peering inside research networks: Some observations on the effect of the intensity of collaboration on the variability of research quality. Research Policy 34, 784-794.

Rigby, J., 2009. Comparing the scientific quality achieved by funding instruments for single grant holders and for collaborative networks within a research system: Some observations. Scientometrics 78, 145–164.

Rogers, J.D., Bozeman, B., Chompalov, I., 2001. Obstacles and opportunities in the application of network analysis to the evaluation of R&D. Research Evaluation 10, 161-172.

Russell, J.M., del Río, J.A., Cortés, H.D., 2007. . Highly visible science: A look at tree decades of research from Argentina, Brazil, Mexico and Spain. INCI 32, p.629-634.

Russell, J.M., 1995. The increasing role of international cooperation in science and technology research in Mexico. Scientometrics 34, 45-61.

Salant, P., Dillman, D.A., 1994. How to conduct your own survey. John Wiley and Sons, New York, NY.

Scandura, T.A., 1992. Mentorship and career mobility: An empirical investigation. Journal of Organizational Behavior 13, 169-174.

Scott, J., 1988. Social Network Analysis. Sociology 22, 109-127.

Scott, J., 2000. Social network analysis: a handbook. Sage Publications Ltd, London, UK.

Small, H.G., 1973. Co-citation in scientific literature: A new measure of the relationship between publications. Journal of the American Society for Information Science 24, 265–269.

Small, H.G., 1977. Co-citation model of a scientific specialty: – a longitudinal study of collagen research. Social Studies of Science 7, 139–166.

Small, H.G., 1978. Cited documents as concept symbols. Social Studies of Science 8, 327-340.

Sargent, LD & Waters, LE 2004, 'Careers and academic research collaborations: An inductive process framework for understanding successful collaborations', Journal of Vocational Behavior, vol. 64, no. 2, pp. 308–319.

Sonnenwald, D.H., 2007. Scientific collaboration. Annual Review of Information Science and Technology 41, 643–681.

Sorensen, C 2003, ‘Scandinavian versus uk research: The importance of institutional Context’, Scandinavian Journal of Information Systems, vol. 15, no. 1, pp .

Spirin, V. and Mirny, L.A., 2003. Protein complexes and functional modules in molecular networks. Proceeding of the National Academy of Science 100, 12123–12128.

Stokols, D Misra, S Moser, R Hall, K & Taylor, B 2008, ‘The ecology of team science - understanding contextual influences on transdisciplinary collaboration’, American Journal of Preventive Medicine, vol. 35, no. 2, pp. S96–S115.

Thomson-ISI, 2003. National Science Indicators database – Deluxe version. Thomson Research, Philadelphia, PA.

Thomson Scientific, 2004. National Citation Report 1980-2003: customized database for the Mexican Council for Science and Technology (CONACYT). Thomson Research, Philadelphia, PA.

Tenenbaum, H.R., Crosby, F.J., Gliner, M.D., 2001. Mentoring Relationships in Graduate School. Journal of Vocational Behavior 59, 326–341.

van Raan, A.F.J., 2000. R&D evaluation at the beginning of the new century. Research Evaluation 9, 81-86.

van Raan, A.F.J., 2004. Measuring Science Capita Selecta of Current Main Issues, in: Moed, H.F., Glänzel, W., and Schmoch, U. (Eds), Handbook of Quantitative Science and Technology Research. Kluwer Academic Publishers, Dordrecht, pp.19-50.

van Raan, A.F.J., 2005. Fatal Attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods. Scientometrics 62, 133-143.

Veloso, F.M., Reyes-Gonzalez, L., Gonzalez-Brambila, C.N., 2005. The Scientific Impact of Developing Nations. Working Paper. Carnegie Mellon University, Pittsburgh, PA.

Verspagen, B., Werker, C., 2004. Keith Pavitt and the Invisible College of the Economics of Technology and Innovation. Research Policy 33, 1419–1431.

Wagner, C.S., Brahmakulam, I., Jackson, B., Wong, A., & Yoda, T., 2001. Science and technology collaboration: Building capacity in developing countries? (Vol. MR-1357.0-WB). RAND, Santa Monica, CA.

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

Wagner, C.S., 2008. The New Invisible College. Brookings Institution Press, Washington, DC.

Waldinger, F., 2010. Quality Matters: The Expulsion of Professors and the Consequences for PhD Student Outcomes in Nazi Germany. Journal of Political Economy 118, 787-831.

Wasserman, S., Faust, K., 1994. Social Network Analysis: Methods and Applications. Cambridge University Press, Cambridge, UK.

Williamson, I.O., Cable, D.M., 2003. Predicting early career research productivity: the case of management faculty. Journal of Organizational Behavior 24, 25-44.

White, H.D., McCain, K.W., 1998. Visualizing a discipline: An author co-citation analysis of information science, 1972–1995. Journal of the American Society for Information Science 49, 327–355.

Woodman, R.W., Sawyer, J.E., Griffin, R.W., 1993. Toward a theory of organizational creativity. Academy of Management Review 18, 293–321.

Wray, KB 2006, ‘Scientific authorship in the age of collaborative research’, Studies In History and Philosophy of Science Part A, vol. 37, no. 3, pp. 505–514.

Wuchty, S., Jones, B.F., Uzzi, B., 2007. The Increasing Dominance of Teams in Production of Knowledge. Science 316, 1036-1039.

Zucker, L.G., Darby, M.R., Brewer, M.B., 1998. Intellectual Human Capital and the Birth of U.S. Biotechnology Enterprises. The American Economic Review 88, 290-306.

Zucker, L.G., Darby, M.R., 2007. Star Scientists, Innovation and Regional and National Immigration. National Bureau of Economic Research, NBER working paper No. 13547 (downloaded on 26 February 2012 http://www.nber.org/papers/w13547.pdf)

Zucker, L.G., Darby, M.R., 2010. Star innovators and high-tech firm entry. National Bureau of Economic Research, NBER Conference Papers (downloaded on 26 February 2012 http://www.nber.org/public_html/confer/2010/SI2010/PRB/Zucker_Darby.pdf)

Zuckerman, H.A., 1967. Nobel Laureates in Science: Patterns of Productivity, Collaboration, and Authorship. American Sociological Review 37, 391-403.

Zuckerman, H.A., 1987. Forward, in Jackson D.N., Rushton J.P. (Eds.), Scientific Excellence, Origins and Assessment. Sage, Newbury Park.

Appendix A. Institutional Codes

A.1. Overview

In this section we provide a list of all the Mexican institution and their respective acronyms used in chapter 2, table A.1 provides this list.

Table A.1. Institutional Codes

Acronym	Institution Name
UNAM	Universidad Nacional Autónoma de México,
CINVESTAV	Centro de Investigación y de Estudios Avanzados,
UAM I	Universidad Autónoma Metropolitana – Iztapalapa
BUAP	Benemérita Universidad Autónoma de Puebla
CIO	Centro de Investigación en Óptica
INAOE	Instituto Nacional de Astrofísica, Óptica y Electrónica
IPN	Instituto Politécnico Nacional
UASLP	Universidad de San Luis Potosí
CICESE	Centro de Investigación Científica y de Educación Superior de Ensenada
UniGuan	Universidad de Guanajuato
UniSon	Universidad de Sonora
ININ	Instituto Nacional de Investigaciones Nucleares
UAM A	Universidad Autónoma Metropolitana – Azcapotzalco
UAEdoMor	Universidad Autónoma del Estado de Morelos
UniGDL	Universidad de Guadalajara
UAZ	Universidad Autónoma
UAQ	Universidad Autónoma
UniMich	Universidad Michoacana de San Nicolás de Hidalgo
IMP	Instituto Mexicano del Petróleo
ITESM	Instituto Tecnológico y de Estudios Superiores de Monterrey
CIMAT	Centro de Investigación en Matemáticas
UDLAP	Universidad de las Américas Puebla
UAEdoMex	Universidad Autónoma del Estado de México
UANL	Universidad Autónoma de Nuevo León
UABC	Universidad Autónoma de Baja California

Appendix B. Sensitivity Analysis

B.1. Overview

In this section we present the sensitivity analysis of the OLS and LOGIT models (of tables 3.8, 3.9, 3.11 and 3.12) using a two standard deviation definition (2SDV), instead of a one standard deviation (1SDV), for the star scientists definition. The impact of entering the systems with a star and a top-PI that has been defend with a more stringent measure (i.e. 2SDV) is higher.

B.2. Regression Models

Table B.1 shows the productivity increase using the second definition for eminent scientists based on productivity (i.e. number of articles). Researchers that entered the system with star-2SDV were on average 37% more productive than the ones that did it with a star-1SDV. And the ones that did it with a top-PI 2SDV were on average 27% more productive than the ones that did it with a top-PI 1SDV.

Table B.1. Productivity increase by type of entry, articles per year, 2SDV, 1984-2001

Type of Entry (Std. Err.) [Total Effect*]	AIb-01	AI-02	AIb-03	AI-04	AIb-05	AIIb-01	AIIb-02	AIIb-03	AIIb-04
Star**, articles	0.379 [c] (0.041) [32%]					0.352 [c] (0.041) [38%]	0.339 [c] (0.042) [36%]	0.352 [c] (0.041) [38%]	0.346 [c] (0.042) [37%]
coAU in RG, articles		0.136 [c] (0.035) [15%]					0.061 [a] (0.036) [7%]		
coAU in top RG**, articles			0.426 [c] (0.085) [46%]			0.287 [c] (0.085) [31%]	0.270 [c] (0.085) [29%]		
coAU-PI of RG, articles				0.098 [c] (0.031) [11%]					0.021 (0.031) [NA]
coAU-PI of top RG**, articles					0.569 [c] (0.111) [64%]			0.444 [c] (0.111) [48%]	0.435 [c] (0.112) [47%]

* Total Effect = coefficient divided by the sample's average
** Star and top RG defined based on the average plus two standard deviations
[a] 10% confidence level, [b] 5% confidence level, [c] 1% confidence level
NA coefficient is not significant

Table B.2 shows the impact increase using the second definition for eminent scientists based on impact (i.e. number of citations). Researchers that entered the system with star-2SDV received on average 53% more citations than the ones that did it with a star-1SDV. And the ones that did it with a top-PI 2SDV had 102% more citations than the ones that did it with a top-PI 1SDV.

Table B.2. Productivity increase by type of entry, citations per year, 2SDV, 1984-2001

Type of Entry (Std. Err.) [Total Effect*]	CIb-01	CI-02	CIb-03	CI-04	CIb-05	CIIb-01	CIIb-02	CIIb-03	CIIb-04
Star**, citations	4.356 [c] (0.258) [233%]					4.364 [c] (0.268) [234%]	4.291 [c] (0.268) [230%]	4.327 [c] (0.268) [232%]	4.285 [c] (0.270) [230%]
coAU in RG, citations		0.814 [c] (0.176) [41%]					0.499 [a] (0.168) [27%]		
coAU in top RG**, citations			1.979 [c] (0.484) [106%]			-0.051 (0.474) [NA]	-0.172 (0.475) [NA]		
coAU-PI of RG, citations				0.536 [c] (0.152) [27%]					0.200 (0.146) [NA]
coAU-PI of top RG**, citations					3.063 [c] (0.655) [164%]			0.257 (0.644) [NA]	0.190 (0.646) [NA]

* Total Effect = coefficient divided by the sample's average
** Star and top RG defined based on the average plus two standard deviations
[a] 10% confidence level, [b] 5% confidence level, [c] 1% confidence level
NA coefficient is not significant

Table B.3 shows the likelihood of becoming a leading researcher using the second definition for eminent scientists based on productivity (i.e. number of articles). Researchers that entered the system with star-2SDV were on average 74% more likely of becoming a leading scientist than the ones that did it with a star-1SDV. And the ones that did it with a top-PI 2SDV were on average 69% more likely than the ones that did it with a top-PI 1SDV definition.

Table B.3. Likelihood of becoming a Star by type of entry, articles per year, 2SDV, 1984-2001

Type of Entry (Std. Err.)	AIVb-01	AIV-02	AIVb-03	AIV-04	AIVb-05	AVb-01	AVb-02	AVb-03	AVb-04
Star*,	4.872 ^{c}					4.452 ^{c}	4.047 ^{c}	4.299 ^{c}	4.164 ^{c}
articles	(0.982)					(0.922)	(0.859)	(0.889)	(0.907)
coAU in RG,		1.491 ^{b}					1.533		
articles		(0.249)					(0.421)		
coAU in top			4.213 ^{c}			2.067 ^{a}	1.943 ^{a}		
RG*, articles			(1.568)			(0.815)	(0.764)		
coAU-PI of RG,				1.519 ^{c}					1.099
articles				(0.203)					(0.231)
coAU-PI of top					8.637 ^{c}			3.675 ^{c}	3.584 ^{c}
RG*, articles					(3.592)			(1.803)	(1.767)

* Star and top RG defined based on the average plus two standard deviations
^{a} 10% confidence level, ^{b} 5% confidence level, ^{c} 1% confidence level

Table B.4 shows the likelihood of becoming a leading researcher using the second definition for eminent scientists based on impact (i.e. number of citations). Researchers that entered the system with star-2SDV were on average 16% more likely of becoming a leading scientist than the ones that did it with a star-1SDV. And the ones that did it with a top-PI 2SDV were 77 % more likely than the ones that did it with a top-PI 1SDV definition.

Table B.4. Likelihood of becoming a Star by type of entry, citations per year, 2SDV, 1984-2001

Type of Entry (Std. Err.)	CIVb-01	CIV-02	CIVb-03	CIV-04	CIVb-05	CVb-01	CVb-02	CVb-03	CVb-04
Star*,	8.971 ^{c}					8.339 ^{c}	7.946 ^{c}	8.491 ^{c}	8.461 ^{c}
citations	(1.795)					(1.226)	(1.649)	(1.756)	(1.785)
coAU in RG,		1.865 ^{c}					1.386		
citations		(0.383)					(0.299)		
coAU in top			4.845 ^{c}			1.229	1.673		
RG*, citations			(1.791)			(0.708)	(0.666)		
coAU-PI of RG,				1.444 ^{b}					1.014
citations				(0.226)					(0.171)
coAU-PI of top					7.317 ^{c}			1.755	1.747
RG*, citations					(3.658)			(0.951)	(0.952)

* Star and top RG defined based on the average plus two standard deviations
^{a} 10% confidence level, ^{b} 5% confidence level, ^{c} 1% confidence level

Appendix C. Questionnaire

C.1. Overview

In this section we present the questionnaire we used to collect the information from Mexican Physicists presented in chapter four of this thesis. Carnegie Mellon University's Institutional Review Board (IRB) approved this instrument and the research protocol of this study.

C.2. Questionnaire

This questionnaire is divided in the following sections:

- Section 1 – Personal Data.
- Section 2 – Entry to the research system.
- Section 3 – Influence and impact of different research environments in the development of new researchers.

N.B.: In section three open-ended questions will be asked; please bare in mind this so you do not give information of someone else which is both identifiable and private/sensitive.
If you have any questions about this survey (or my research) do not hesitate to contact me at lreyesgo@andrew.cmu.edu or give me a call +52 (55) 8421-6910.

Section 1 – Personal Data

1.1 Age: ____________ Sex: _____________
1.2 Research Area: ____
1.3 Type of Research (Theoretical/experimental): ______________________________________
1.4 Maximum level of study: ____
1.5 Starting year for maximum level of study: _______________________________________
1.6 Finishing year for maximum level of study: _____________________________________
1.7 Institution where the maximum level of study was obtained: _______________________________
1.8 Department/School/Faculty where the maximum level of study was obtained: ___________________
1.9 Current institutional affiliation: ____
1.10 Current departmental/School/Faculty affiliation: ____
1.11 Seniority at current institution/department: _________________________
1.13 Category within the (Mexican) National Researcher System: ____________________________
1.14 Considering that an **<u>ISI article</u>** is a manuscript that has been published in a journal indexed by the Institute for Scientific Information, please indicate the number of manuscripts you have published and an approximate number of citations you have received for those publications:

	Number of articles	**Approximate number citations**
ISI Articles		
Other type of <u>articles</u>		
Books		

1.16 Within academia, do you consider that you play a more prominent role than other researchers in terms of the number of:

- Published articles (YES/NO): ______________
- Received citations (YES/NO): ______________

Section 2 –Entry to the research system

2.1 Considering that entry to the system is when you published your first article or book by yourself or with one or more authors, in which year you entered your first manuscript, or entered the system? ______

2.2 Institution you were affiliated when you entered the system: ________________________________

2.3 Department/School/Faculty you were affiliated when you entered the system: ____________________

2.4 Highest degree obtained when you entered the system: ________________________________

2.5 Was your first publication an ISI article (YES/NO): ________________________

2.6 In which year you published your first ISI article? ____________________

2.7 Please indicate under which collaboration scheme you published your first five articles (or less if you have published a lower number):

Collaboration scheme*	**Number of articles**
Individual publications, where you are the only author	
Publications with only one coauthor	
Publications with two or more coauthors up to six coauthors	
Publications with seven or more coauthors	

*Mutually exclusive categories

<u>In case your first five manuscripts were individual publications please continue in section 3.</u>

2.8 Do you consider that **any of your coauthors of your first five publications** (or less if you have published a lower number) **<u>was playing at the time of publishing these manuscripts</u>** a more prominent role in academia than other researchers in terms of:

- Published articles (YES/NO): ______________
- Received citations (YES/NO): ______________

2.9 Considering that a **Research Groups (RG)** is a group of people that collaborates repetitively in scientific research and publishes the results of these activities in articles or books, please indicate for your first five publications (or less if you have published a lower number) the number of articles that were developed/elaborated within a RG: __________

<u>In case your first manuscripts were not developed/elaborated within a RG please continue in section 3.</u>

2.10 Your first five manuscripts (or less if you have published a lower number) were developed/elaborated in the same RG (YES/NO): _______

2.11 What was the average size of these RGs (number of researchers)? _____________

2.12 What was the maximum size of these RGs (number of researchers)? _____________

2.13 Do you consider that **any of the RG where you developed/elaborated your first five publications** (or less if you have published a lower number) **<u>was playing at the time of publishing these manuscripts</u>** a more prominent role in academia than other RG in terms of:

- Published articles (YES/NO): ______________
- Received citations (YES/NO): ______________

2.14 Considering that a **Principal Investigator of a research group (PI-RG)** (defined in question 2.9) is the scientific leader of the group, please indicate for your first five publications (or less if you have published a lower number) the number of articles that were developed/elaborated with the PI-RG: ______

<u>In case your first manuscripts were not developed/elaborated with the PI-RG please continue in section 3.</u>

2.15 Do you consider that **any of PIs you have collaborated on your first five publications** (or less if you have published a lower number) **<u>was playing at the time of publishing these manuscripts</u>** a more prominent role in academia than other researchers in terms of:

- Published articles (YES/NO): ______________
- Received citations (YES/NO): ______________

Section 3 – Influence and impact of different research environments in the development of new researchers.

3.1 Considering that

- A FORMAL COLLABORATION is when you are invited to collaborate after participating in a competitive process at a research institution/program and being admitted to it (like admission to a PhD program or post doctoral position),
- An INFORMAL COLLABORATION is when you were invited to collaborate without such competitive process, and
- A COLLABORATION YOU LOOKED FOR is when you actively searched for that collaboration by yourself;

please indicate for your first five publications (or less if you have published a lower number) and the following research environments how you started to collaborate within a particular environment.
Chose all the options that apply in your case, you can chose more than one option per row.
If you have never collaborated in such environment choose N/A (no available):

esearch environment	How you start your collaborations? Through a FORMAL collaboration process:	Through a INFORMAL collaboration process:	You looked for the collaboration:	Other form of collaboration, please specify in the next question:	N/A
ıblish individually:					
ollaborate and publish with only one coauthor:					
ollaborate and publish with two or more coauthors up to x coauthors:					
ollaborate and publish with seven or more coauthors:					
ollaborate and publish with a researcher that on average ıd at the time this collaboration started a higher:					
Number of articles than the majority of scientists in the same field of research:					
Number of citations than the majority of scientists in the same field of research:					
ollaborate and publish in a research group (RG):					
ollaborate and publish within a RG that on a average ıd at the time this collaboration started a higher:					
Number of articles than the majority of groups in the same field of research:					
Number of citations than the majority of groups in the same field of research:					
ollaborate and publish with a principal investigator of a search group (PI-RG):					
ollaborate and publish with a PI-RG that on average had the time this collaboration started a higher:					
Number of articles than the majority of scientists in the same field of research:					
Number of citations than the majority of scientists in the same field of research:					
ther research environment, specify:					

3.2 In case you chose OTHER form of collaboration to the previous question please specify how you started this collaboration?

ch environment	How you started this collaboration?

3.3 Based on a scale of 1 to 4 (where 1 is “no impact” and 4 is “high impact”) please indicate the impact the following research environments have had in your scientific career.
You can only chose ONE OPTION per row.
If you have never collaborated in such environment choose N/A (no available):

Research environment	1 No Impact	2 Low Impact	3 Medium Impact	4 High Impact	N.A.
Publish individually:					
Collaborate and publish with only one coauthor:					
Collaborate and publish with one or more coauthors up to six coauthors:					
Collaborate and publish with seven or more coauthors:					
Collaborate and publish with a researcher that on average had at the time this collaboration started a higher:					
Number of articles than the majority of scientists in the same field of research:					
Number of citations than the majority of scientists in the same field of research:					
Collaborate and publish in a research group (RG):					
Collaborate and publish in a RG that on a average had at the time this collaboration started a higher:					
Number of articles than the majority of groups in the same field of research:					
Number of citations than the majority of groups in the same field of research:					
Collaborate and publish with a principal investigator of a research group (PI-RG):					
Collaborate and publish with a PI-RG that on average had at the time this collaboration started a higher:					
Number of articles than the majority of scientists in the same field of research:					
Number of citations than the majority of scientists in the same field of research:					
Other research environment, specify:					

In the remaining part of this section, open-ended questions will be asked; please bare in mind this so you do not give information of someone else which is both identifiable and private/sensitive.

3.4 In terms of opportunities, what opportunities these research environments

a) Publish individually:
b) Collaborate and publish with only one coauthor:
c) Collaborate and publish with one or more coauthors:
d) Collaborate and publish with a researcher that on average had at the time this collaboration started a higher a number of articles/citations than the majority of scientists in the same field of research:
e) Collaborate and publish in a research group (RG):
f) Collaborate and publish in a RG that on a average had at the time this collaboration started a higher number of articles/citations than the majority of groups in the same field of research:
g) Collaborate and publish with a principal investigator of a research group (PI-RG):
h) Collaborate and publish with a PI-RG that on average had at the time this collaboration started a higher number of articles/citations than the majority of scientists in the same field of research:
i) Other research environment, specify:

have opened to you? (e.g. access of financial resources; qualified personnel' specialized labs; opportunities to collaborate people at other research institutions, countries; being exposed to new ideas; etc.

Research environment(s) with the highest impact on your scientific career	What opportunities these research environments have opened to you?

3.5 In terms of learning, what have you learned from these research environments (discussed previously)? (e.g. new ideas; different research techniques; write research proposals; publish ISI articles; teach or advise other scientists; collaborate with other researchers; etc.):

Research environment(s) with the highest impact on your scientific career	What have you learned from these research environments?

3.6 Is there another situation or setup that had influence positively or negatively the development of your scientific career. If this is the case pleas indicate, what has been this situation, the impact of it and what have you learned from it?

Situation	Impact	What have you learned from this situation?

Are you willing to have a more in-depth conversation with us about this topic and answer some additional questions (YES/NO): ____________________
If you answered positively the previous question please give us your contact information:
Name: ____
Telephone: _____________________________________
e-mail: ____

The questionnaire ends here
Thank you for you help