Measuring Intellectual Diversity in Groups

by

Alexander Graham

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Master of Arts

in

Sociology

Waterloo, Ontario, Canada, 2017

Author Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

I answer questions regarding the relationship between the intellectual diversity of an individual's collaborations and their impact and productivity. I propose two new measures of intellectual diversity: coauthor diversity score and group intellectual diversity. Coauthor diversity score measures how different an author is from their immediate collaborators. Group intellectual diversity measures how diverse a group is as a composite of each members difference to every other member. These measures allow me to test hypotheses regarding the relationship between intellectual diversity and career outcomes. I conduct two case studies, using data from the Web of Science database, to examine the measures in biomechanical modelling and nanotechnology. I bootstrapped negative binomial regressions with squared interaction terms for diversity measures to test for relationships between individual and group diversities against career citations. Increasing diversity is associated with more citations until an maximum after which citations exhibit negative returns as diversity increases. This behaviour is consistent in both case studies. I use a bootstrapped linear regression to show that individual and group diversities have a significant effect on publication rates in biomechanical modelling, but not in nanotechnology.

Acknowledgments

It is truly impossible to fully acknowledge everyone who has helped me get to this point. I will, however, attempt the impossible. Firstly, I would like to thank my advisor, Dr. John McLevey, who opened my eyes to sociology and network analysis. When I got an email saying that the new professor in Knowledge Integration was looking for research assistants, I sent in my resume, not thinking much of it. I had never taken a sociology course. I had no idea what social network analysis was. When I got an email asking if we could meet, it changed the course of my life. I quickly fell in love with network analysis, eventually leading to the point where I decided that I would pursue a Master's degree in sociology. Along the path, I could not have asked for a more supportive, enthusiastic, and knowledgeable advisor. I can say without a shadow of a doubt that if I had not met John, my life would be dramatically different, and I suspect rather less fun.

I would like to thank Dr. Martin Cooke. He has not only been incredibly helpful as the Associate Chair for Graduate studies, but gave me an opportunity to collaborate with him in an area outside my immediate research. Working with him has really helped broaden my perspective about the scope of what I have to offer and helped me think about new ways that I can leverage my expertise to solve problems outside the sociology of science. His comments during the process of producing this thesis have been instrumental to its formation. He pushed me to get my head out of the clouds and produce concrete answers

to the important questions: "What does this mean and who cares?" Without his questions, I would not have the answers that I have today.

I would like to thank Dr. Peter Carrington. His social network analysis course was foundational to how I have thought about networks ever since I took it. Not only that, it introduced me to important people who have featured heavily in my life since then, and without whom my life would certainly be poorer. Every time Peter has offered his thoughts and advice, it has been exactly what I needed. Whether it be with regards to navigating the murky waters of academic grant writing or producing actually meaningful work, Peter has always been incredibly helpful. I spent months working on refining an idea I'd been toying with for about a year. Peter asked me a single question about whether it makes sense to focus on artifacts or people when my research is on creativity and collaboration. It helped me realize all of the fundamental issues that were plaguing me and how I could quite easily resolve them by changing how I looked at the problem. My data, methods, and analysis changed radically, but all for the better, becoming cleaner, easier to interpret, and better able to answer the questions I was actually asking. Without Peter, this thesis would look completely different, and would most certainly have been much worse.

Dr. Owen Gallupe, although not an official member of my thesis committee, often made himself available for conversations about modeling, and was the one who suggested including Tobit models in this thesis. I am grateful to him for always being available for advice, especially when I was trying to wrap my head around interaction terms in non-linear models.

I would like to thank the members of Dr. McLevey's *NetLab* at the University of Waterloo. Reid McIlroy-Young was instrumental to the creation of *metaknowledge*, a software package absolutely integral to my research. Without Reid and Dr. McLevey's work, what

appraisal about balancing work, life, and self-care. I don't know anyone else who keeps it real like Brittany. Sascha Lecours is a bright star of laughter in my life. When he isn't giving me some of the best real-life advice I've ever received, from any source, he is testing the human limits for laughter with me as one of his test subjects. Before I met Sascha, I never thought that a dramatic re-telling of the history of lobster as a luxury item would be, not just interesting, but one of my favourite stories ever. Nor would I have given much thought to how one should go about balancing renting vs buying a house, or investment options. John Petersons has always been one of my most important pillars of support. On many occasions, I've been completely stumped and we've shared our puzzles. When all else fails, I can always trust John to be there to come up with a super simple solution, or the right question that eventually leads to the solution. His patience while putting up with me will never be forgotten.

I would like to thank my family. Their constant support has helped me become the person that I am now. My parents have worked very hard to give me an incredible start in life and they have constantly pushed me to be the very best. I will never forget my father's advice to me: "Find something you love, become the best at it, and then find someone who will pay you to do it." So far, I've got the first part down. This thesis is one more step on the second part. My grandfather has been a huge factor in my life and my view of the world is so much broader because of him.

Finally, I would like to thank Claire Gallant, my fiancée. Thanking her properly would be a thesis in itself, complete with table of contents and appendices. In lieu of that, I will instead settle with saying the Claire is the kindest, most warm-hearted, and hardest working person I know. Claire has stayed up at night to provide moral support and tea when deadlines loom on the horizon. She has listened and joined in with every idea, complaint,

celebration, set-back, and moment of triumph. She has helped me keep me organized and keep track of my life outside academia when it feels like my thesis is the only thing that exists. Without her, I would never have had the courage to apply to a Master's program. Without her, I would not be half the person that I am today. Her never-ending pride in me has been a constant source of motivation to "get shit done," as she would say.

Table of Contents

List of Figures

List of Tables

1

Introduction

Researchers in the sociology of science, organizational management, and social psychology have been examining predictors of innovation and productivity in the workplace for decades (Hülsheger, Anderson, and Salgado 2009; Kern 2011). With the shift towards knowledge-based economies, the ability to innovate is of paramount importance to companies and governments who have devoted billions of dollars and other resources to trying to improve the quality of decision-making processes by increasing diversity (Phillips 2014).

With all the time and money devoted to increasing diversity, it is important to understand what forms of diversity matter and the conditions and outcomes for which that diversity is helpful, unhelpful, or has no effect. Hülsheger, Anderson, and Salgado (2009) note that different forms of diversity have different effects in different contexts. For example, this means that the positive effect of gender diversity in firm performance may not be the same if one is looking at the publishing speed of a lab group. Diversity is too general a concept to expect it to act consistently in all contexts. The literature on diversity and collaboration has focused primarily on cohesive groups with clearly defined boundaries and high-level measures of diversity for groups. This has been to the detriment of the study of

diffused groups and the effects of individual diversity with respect to the rest of the group.

Given how general the concept of diversity is, and how many different outcomes for group collaborations one might be interested in, it would be infeasible to address all forms of diversity and outcomes. For the scope of this analysis, I will focus on only one form of diversity and its association with two outcomes. I will be examining the relationships between a diversity of intellectual capital, impact, and productivity. To address the afore-mentioned gaps in the literature, I focus on *intellectual* diversity as a means of highlighting graduated individual differences between direct collaborators and loose groups of coau-thors in publishing in academic fields[1] to represent informal groups consisting of direct and indirect collaborations.

To perform my analysis, I introduce two measures of intellectual diversity: coauthor diversity score and group intellectual diversity. An *individual's* coauthor diversity score (CDS) is a summary measure of how different someone is from the total of their formal collaborators. For a *collection of authors*, group intellectual diversity is a summary measure of how different each member is to every other member. I focus on these measures to examine questions about diversity occurring within direct and indirect collaborations to identify possible differences and similarities.

---

[1]In this thesis, I will be using several terms that have ambiguous meanings. I will endeavour to use them as precisely and consistently as possible in the following ways: 1) I use "discipline" to refer large bodies of research roughly akin, but not necessarily identical, to university departments: Sociology, Biomechanics, Physics, etc. 2) I use "field" to refer to specific areas of research within a discipline: Biomechanical modelling, nanotechnology, environmental GIS, etc. 3) I use "group" to refer to the clusters of authors identified by modularity class in a field's coauthorship network. 4) I use "researchers" and "authors" interchangeably.

2

Motivation

In this section, I will motivate my research with an examination of two particularly salient examples of informal collaborative groups. The first is the Linux kernel, a massive Free and Open-Source Software (FOSS) project that underlies much of the technology that we rely upon. Individuals working on the same files may be working for competing companies in different countries at completely different points in time. They may never even communicate with one another. Meanwhile communication and organization is managed through hundreds of mailing lists, some with hundreds of messages every day. The second example is the European Organization for Nuclear Research (CERN). While membership at CERN may be clear, formal collaboration happens at level of research teams: not all members of CERN collaborate with each other, despite sharing the same space, equipment, and field of study. Collaboration within both of these contexts involve both formal collaborations, where individuals with different skill sets will work together to solve a problem, and informal collaboration, where individuals will observe the efforts of others and learn from them or build on their work. In this thesis, I am focused on measuring the effects of diversity within these two kinds of collaborations.

Large, diffuse projects have several characteristics that distinguish them from the cohesive or monolithic groups commonly examined in the literature. Goals are broader and less specific. Membership is less strictly regimented. Organization is more decentralized. Collaboration comes in varying degrees and flavours. Formal collaboration is often specific to a particular subset of the project. However, a less formal collaboration also takes place where authors learn from and build on the work of others without directly interacting with them.

These distinctions are important because nearly every paper looking at diversity has examined clearly defined groups of people: research labs, project teams, and boards of directors (Dayan, Elbanna, and Di Benedetto 2012; Kern 2011; Uzzi et al. 2013). No matter the level of analysis, groups had clear boundaries and goals. Even groups that are far too large and diffuse for everyone to work together, like companies with hundreds or thousands of employees, are treated as monolithic entities where diversity and collaboration reach all members equally. While the examination of diversity and collaboration in small formal groups is important, they are not the only kinds of groups where diversity and collaboration take place. I will focus on large and diffused groups that contain both formal and informal collaboration because some of the leading edge of science and software development occurs in such places. Two such groups are the developers of the Linux kernel and the scientists at CERN.

The Linux kernel is the underlying code that manages the interaction between an operating system's applications and the physical hardware that comprises the computer (Foundation 2013). Specifically, it manages these interactions for every distribution of the Linux operating system, of which there is no official count. However, Distrowatch.com indicates that there are at least 863 distributions of Linux (Distrowatch 2017).

The Linux kernel is an enormous Free/Open-Source Software (FOSS) project with regular 2-3 month releases, each with over 10,000 patches, containing the work of more than 1100 developers from over 225 corporations (Foundation 2013). Between 2005 and 2013, almost 10,000 individual developers from over 1000 companies contributed to the kernel (Foundation 2013). At over 13 million lines of code, the Linux kernel is among the largest open source projects in existence.

Given the sheer scope of the project, communication is key to organizing the process. The main mailing list, the Linux Kernel Mailing List, was seeing 440 emails per day in 2010, and a number that has been consistently growing larger since 1995 (Menzies, Williams, and Zimmermann 2016). However, there are 200 more mailing lists devoted to sub-systems of Linux, some seeing as many as 30 messages per day.

Started as a personal project by Linux Torvalds in 1991, Linux distributions have become fundamental to the current state of technology. It runs not only on computers, but smart phones, routers, web servers, supercomputers, TVs, refrigerators, tablets, stock markets, nuclear submarines, and power stations, to name a few (The Linux Foundation 2017). As of 2013, 81% of smartphones ran on Linux. In 2011, 64% of servers and 92% of supercomputers ran on Linux (The Linux Foundation 2017). Linux is so important to so many different technologies that the kernel is being actively developed by companies that are nominally in competition with each other: Red Hat and Oracle, Fujitsu and HP, AMD and Intel, and so on (McLevey and Graham n.d.).

In a similarly massive scale in terms of numbers of collaborators, but far larger in budget than the Linux kernel, the European Organization for Nuclear Research, commonly known as CERN, consists 20 European Union Member States with 2531 internally funded staff, 13,794 fellows, associates, students, users, and apprentices. These members come from

approximately 608 institutes and universities from around the world as of 2015 (European Organization for Nuclear Research 2015).

Initially founded in 1952 to establish a world-class European research organization in fundamental physics, it has grown to include many areas of particle physics as our understanding has deepened (CERN 2017a). By bringing together the resources of its member states and organizations, CERN is able to construct equipment that would be far too expensive for any one group alone, such as the Large Hadron Collider (LHC), a 27-kilometre ring of superconducting magnets, buried under the French-Swiss border, that must be cooled to -271.3°C (CERN 2017b). Constructing the LHC cost approximately $4.74 billion dollars, and operating costs run at about $1 billion dollars per year (Knapp 2012). Since these costs would be inordinately high for a single nation, the members of CERN have pooled their resources and collaborated to create the organization, facilities, and equipment that facilitates high-energy particle physics that could not otherwise be done.

Just as the construction and funding of the LHC was a massive work of collaboration, the projects conducted under CERN's umbrella are likewise large. One paper coming out of CERN set the record for the highest number of authors, 5,154; its 33 pages devote 9 pages to the research and 24 pages to listing the authors and their institutions (Castelvecchi 2015).

While the contexts, motivations, skills, and actors in these two examples are very different, they share several important similarities that highlight the motivation behind my proposed measures for diversity. They are projects too massive for every contributor to meaningfully collaborate with every other contributor. Organization of individual projects is impossible for a centralized decision-making body, so projects are decided at a much

lower level while organization serves to ensure cohesiveness, allocate resources, and re-solve conflicts. Even after the dividing the project into smaller sub-units, such as papers or patches, the number of contributors is still enormous (European Organization for Nu-clear Research 2015; The Linux Foundation 2017). They require contributors with a wide variety of backgrounds and expertise for success. Group membership boundaries are very porous. Individual sub-projects share resources and an intellectual space, either physical or digital. Sub-projects may or may not be directly related to each other, but they all con-tribute in some way to the shared goals of the project. These aspects are important because they indicate the presence of both direct collaboration to complete sub-projects and indirect collaboration to further the overall goals of the project.

I argue that the current literature on group projects does not adequately describe the conditions under which large, uncohesive projects like CERN or the Linux Kernel operate. I further argue that CERN is not radically different than other academic fields in principle, but rather, it is a more visible and scaled up example of academic research in scientific fields. Its projects are larger, more expensive, and higher profile, but much of the same mechanisms are in effect in other fields at a smaller scale. CERN, at its core, is a resource-sharing space for the study of particle physics (CERN 2017a). While many of the experi-ments being done at CERN are related to each other, and often multiple research teams will use data from the same experiments, its major difference from the normal science of less expensive fields of study is that researchers are working in the same geographical location and using the same equipment. Even this is not so different from how a university operates, where different researchers and their teams may use the same equipment and spaces for different projects in the same general field of study. The distinction is not one of form, but of scale.

This similarity is important because it allows one to study academic fields more generally. This allows researchers to make use of the data, methods, and tools that have already been developed for scientometrics and bibliometrics. While CERN is certainly one of the most recognizable cases, the scale of the research and the number of authors on each paper has consequences that make analysis more difficult. Instead, I will focus on two fields, also in the natural sciences, but more manageable in scope. I will be studying biomechanical modelling and nanotechnology.

In these cases, direct collaboration is quite easy to parse out as coauthorship levels are much more reasonable. Indirect collaboration is a feature of being present and aware of the field. While the contributors of the Linux kernel are working to improve the kernel, everyone at CERN is loosely working to advance the knowledge of particle physics. In both of these cases, members of the field are working together towards the poorly defined goal of "improving the kernel" or "advancing knowledge". One's work is often contingent on other people doing their own jobs well and learning from them, even if not working with them. Owen-Smith and Powell (2004) refers to this as a kind of spillover effect from "leaky" collaborations. Here, I draw a connection to the words of Sir Isaac Newton, "If I have seen further, it is by standing on ye shoulders of giants" (Newton 1676). Such massive projects as the ones I have described are advanced not by individuals, or even a team of individuals, but many groups and individuals all working independently, and yet building on each others'works.

3

Literature Review

The literature on diversity in groups at the very broadest level tends to focus on two major aspects. The first is the form of diversity. The second is the outcome that the authors expect to be associated with the specific form of diversity.

I will examine the prior work that has been done to define and organize diversity. I will then summarize the several major theories and hypothesized mechanisms that suggest *how* diversity may have an effect on group outcomes, focusing on creativity as a mediator of the outcomes that I am interested in. While the outcomes I am focusing on in this paper are impact and productivity, I do not hypothesize a direct relationship between diversity of intellectual capital and these outcomes. Rather, the most common mechanism described in the literature is the effect of diversity on the creative and decision-making processes of group work. Having summarized the literature on diversity and creativity, I will explain how one would expect diversity to then interact with impact and productivity: mediated by its effect on creativity.

Because I am proposing two new measures of diversity, I will summarize previous attempts to measure diversity and explain why they do not meet the needs that my measures

seek to fulfill.

3.1   Defining and Organizing Diversity

The literature on the benefits of diversity on innovation in groups spans decades and disciplines (Hülsheger, Anderson, and Salgado 2009; Phillips 2014). In this section, I will summarize some of the major attempts to define and organize the concept of diversity.

When it comes to solving large and complex problems, diversity of expertise would appear to be superior over homogeneity. This can be seen in how nearly every team for such problems is organized. Designing an airplane takes knowledge of aerodynamics, propulsion, materials, structural mechanics, and electronics. Leading an army takes knowledge of strategy, supply, logistics, geography, and more. Building a house takes knowledge of plumbing, wiring, insulation, structural loads, aesthetics, and more. While there is of course homogeneity within these teams, the existence of these different roles highlights that specific forms of diversity are desired. It is unreasonable and inefficient to expect a single person to acquire the necessary expertise in every necessary area. The logical conclusion is to create teams of individuals who are experts in their own disciplines and synthesize their efforts to create the final product, solution, or plan.

Diversity as a concept is very general. When one talks about diversity, it can easily be a diversity of opinions, ideas, cultures, genders, or socio-economic statuses. While this generality is one of the reasons it is so interesting to study, it also means that one must be especially careful about how one defines and measures diversity.

Researchers have consistently distinguished between two varieties of diversity: informational diversity — education, occupation, project background — and social category or sociodemographic diversity — race, gender, age, socioeconomic status — (Dayan, Ozer,

and Almazrouei 2017). When operationalizing these forms of diversity, especially social category diversity, researchers often rely on single-choice categorical categorizations. For example, someone is a single race, ethnicity, or nationality. While simple and easy to comprehend — two things are either the same or they are different — it misses an opportunity for improving the granularity of our measures of diversity. While applicable to both informational and sociodemographic diversity, this is easier to see when discussing informational diversity. Suppose two people were to attend the same party. The first person stays for the first two hours. The second arrives an hour late, but stays for four hours. If one were to ask them what happened at the party, they would both have some pieces of information that were different, and some that were the same. If one were to ask both of them together what happened, one would expect a fuller image of what happened than if one talked to them individually. This spectrum of diversity is an often-overlooked and under-developed area of research. Some research has gone into dyadic diversity measures, of the kind I have just described, however these are much rarer and their use in empirical studies is still very young (Soós and Kampis 2010).

Sociodemographic diversity, as the name suggests, is diversity measured along demographic axes such as gender, age, race, or socioeconomic status. This particular form of diversity has the most debate about its actual meaning and the mechanisms by which it interacts with group outcomes. Dayan, Ozer, and Almazrouei (2017) suggests that demographic diversity is a kind of essentialist line of reasoning, relying on stereotypes and preconceived notions. Harvey (2013) suggests that racial diversity is a valid proxy for cultural diversity, or a diversity of perspectives and ways of viewing the world, since racial experiences colour and influence our experiences. Both Dayan and Harvey would argue that the important kind of diversity is that which arises from differences in information or

approaches to problem solving. Phillips (2014) argues takes a different approach, arguing that sociodemographic diversity is itself an important feature of diversity because its salience encourages people to make fewer assumptions of shared information, resulting in greater information sharing.

Some researchers identify differences in perspectives, attitudes, and values as important sources of deep level, or psychological, diversity (Harvey 2013; Ziller, Behringer, and Goodchilds 1962). However, these are distinguished from informational diversity, or diversity resulting from such things as differences in education, occupation, or experience, which instead serve as proxies for deep diversity (Mannix and Neale 2005; Milliken and Martins 1996). The distinctions between informational and deep diversity are much finer than the distinction between informational and sociodemographic diversity.

Based on this very rudimentary taxonomy, one might suggest different mechanisms and effects for such forms of diversity as race (sociodemographic), education (informational), and political leanings (deep). Unfortunately, this section of the literature is particularly fractured and inconsistent. No clear consensus seems to have arisen regarding a proper taxonomy of forms of diversity. Furthermore, it remains to be seen whether a rigorous taxonomy of diversity would provide meaningful and useful framing for research.

3.2   Defining Creativity and the Creative Process

Although I intend to analyze the relationships between intellectual diversity and an author's impact and productivity, the proposed mechanism by which diversity would have an effect on these two outcomes is mediated by its effect on the creative process. I will first summarize some of the major theories about what creativity is. I will then explain the creative process and the proposed means by which diversity may help or hinder it.

Creativity is generally thought of as being difficult to define, yet it remains a valued quality in academic thought and work. Mishra et al. (2013) conducted a review of 90 peer-reviewed articles specifically relating to creativity and found that only 38% of them provided a definition of creativity. Medina (2012) argues that not only can creativity not be defined, but one should not even try because it is "a spirit, a gift, a pleasure, and a pain. It is different for you than it is for me, and it catches us all in the most unexpected of places."Howard, Culley, and Dekoninck (2008) identify a group of psychologists that they call romantics: researchers who view creativity as a mysterious, spiritual, or subconscious process, whom they contrast with the more productive and useful non-romantics who view the process of creativity as something concrete that can be studied and learned. There is a thriving field of study which attempts to do exactly what Medina says is impossible: define and even measure creativity. Heinze (2012) argues that creativity is characterized by a particular work being both relevant and novel. Scientific creativity takes the form of work that is scientifically relevant and original (Heinze, 2012). Heinze alludes to the earlier work of Polanyi and Grene (1969) that argues that science is relevant if it is both plausible and of scientific value. By this, he means that a scientific work should be grounded within the established knowledge of the field and provide something to other members of the community. To summarize, a scientific work that is creative should bring new knowledge to the table, that knowledge should make sense given the theories or evidence within the field, and it should provide something of value to the other members of the scientific community.

With a definition of creativity as work that is both relevant and novel, one can begin to examine the forces that drive the expression of creativity. One of the recurring themes in the research on creativity indicates that there is a constant tension between novelty and conformity (Kuhn 1970; Polanyi and Grene 1969; Uzzi et al. 2013). Polanyi and Grene

13

(1969) argues that the need for science to be both plausible and of scientific value acts to impose conformity, whereas the need for works to be novel encourages breaking from that same conformity. Uzzi et al. (2013) attempt to address this exact pressure towards both conformity and novelty in their bibliometric study. They argue that creativity is produced primarily with original combinations of pre-existing knowledge to produce new ideas (Uzzi et al. 2013). As a result, creativity poses the challenge of encouraging scientists to cross disciplinary boundaries to make connections that no one has thought of before (Uzzi et al. 2013). The requirement that the resulting knowledge be both plausible and scientifically relevant further complicates the inter-disciplinary challenges; Uzzi et al. (2013) present the historical example of Sir Isaac Newton presenting his theory of gravity using classic geometry, rather than his newly invented calculus, so that his audience would be able to understand the reasoning and importance of his ideas: the plausibility and scientific value of the theory of gravity. If scientists are becoming more specialized as fields of study grow ever larger, then individual scientists will have greater challenges bridging knowledge silos (Uzzi et al. 2013). The natural extension of this specific challenge is the growth of interdisciplinary teams of scientists. If individual scientists are too specialized to effectively bridge disciplines in a productive manner, then teams of scientists that are experts within their own fields may be brought together to solve a complex problem that would benefit from a diversity of ideas and methods. In fact, there is evidence that this is one of the factors driving the recent increase in co-authorships (Groboljsek et al. 2014).

If one accepts the tension between conformity and novelty as a driving force behind the expression of creativity, one can then attempt to find measures of that expression. Uzzi et al. (2013) provide a strong argument for a specific measure of creativity in a study of meta-data on 17.9 million papers. They determined the probability that two citations from two

14

journals would appear in the same bibliography (Uzzi et al. 2013). Based on this, they created a distribution of co-citations for every pair of citations in a paper's bibliography (Uzzi et al. 2013). From this distribution, they derived two summary statistics: the 50th percentile z-score, which they argued provides an indication of the conventionality of the bulk of co-citations, and the 10th percentile z-score, which provides an indication of the novelty of a paper's co-citations (Uzzi et al. 2013). The authors then split papers into four categories: high novelty and high conventionality, high novelty and low conventionality, low novelty and high conventionality, and low novelty and low conventionality (Uzzi et al. 2013). They then tested whether these categories predicted whether a paper would be a "hit" paper — be in the top 5% of received citations over an 8-year period — and found that papers were hit-status 9.11%, 5.33%, 5.82%, and 2.05% for each of the categories respectively (Uzzi et al. 2013). To summarize, having the bulk of one's citations being highly conventional and introducing a few highly unconventional papers almost doubles the probability of a paper being in the top 5% of received citations over the next best citation pattern. It is important to note that while the improvement is dramatic, relatively speaking, it is still not a sure-fire way to improve citation count. Authors must still synthesize new information from conventional and non-conventional sources in a meaningful and productive form. Thus, one sees evidence that a diversity of ideas, couched in the established wisdom of a scientific field, will on average produce more creative papers.

This understanding of citations being indicative of creativity is not isolated to Uzzi et al. (2013). Collins (1998) argues that intellectual creativity results from combining previous knowledge and that references in a paper are a rough indication of the cultural capital that the paper draws upon. In this case, Collins (1998) is using cultural capital to mean the collection of intellectual and contextual resources that allow scientists working in

a specific field of study to build knowledge and is a necessary requisite for creative work. Like Uzzi et al., Collins (1998) identifies the tension between novelty and relevance, using the example of Einstein's theory of general relativity in Hellenistic Greece; no matter how novel something is and how well it explains evidence, if its audience can not appreciate it, it won't be taken up by the community. While the theory of relativity would have been novel, it would have had no value to the people of the time because it would answer no questions that they were asking and provide no input into discussions they were having.

Kuhn (1970) uses a similar definition creativity for his conception of a paradigm. A paradigm is an achievement that is "sufficiently unprecedented to attract an enduring group of adherents away from competing modes of scientific activity. Simultaneously, it was sufficiently open ended to leave all sorts of problems for the redefined group of practitioners to resolve" (Kuhn 1970). In short, a paradigm is both novel and relevant to the rest of the community by providing a framework for the rest of the community to base their future work on. Paradigms provide a common language for future work while also indicating fruitful avenues of research (Kuhn 1970). In this way, Kuhn's paradigms drive both the conformity and novelty of future works, which would explain their central position in intellectual networks.

Leahey and Moody (2014) further reinforce this definition of creativity, arguing that creativity in academic science is driven when material from one domain "inspires or forces fresh thinking in another", where domains of interest are academic disciplines or subdisciplines. The recombination of previous work in new ideas produce "good ideas, higher quality output, and serve as a foundation for innovation" (Leahey and Moody 2014).

While many of the theories I have just described do a good job of covering the ideas behind creativity, they do less well at describing the nuts and bolts of how creativity works

16

as a process. It is important to understand *how* creativity works if one wants to study it. While I have laid out an outline for what constitutes creativity, how one *produces* creativity is just as important. Just as many authors have attempted to tackle the problem of defining creativity, others have attempted a similar task of defining and categorizing the creative process and its stages. Howard, Culley, and Dekoninck (2008) summarizes a small subset of two different literatures — engineering and cognitive psychology — and in doing so, identifies 23 different formulations of the creative design process ranging from 1967 to 2006. Howard, Culley, and Dekoninck (2008) were attempting to identify an overarching theme to summarize and highlight the often synonymous definitions of the many models that came before, and perhaps ironically, in doing so they added a 24(^{th}). Howard, Culley, and Dekoninck (2008) identifies the following phases:

1. *Analysis Phase*: a problem is identified, its scope is defined, and information is gathered.

2. *Generation Phase*: ideas for possibilities are generated.

3. *Evaluation Phase*: the ideas generated earlier are compared and evaluated to find the one most suitable based on the salient criteria of the circumstances.

4. *Communication / Implementation Phase*: the chosen idea is put into action.

The general idea behind this process is that for a given problem, a group must first identify it, define it, and gather all the relevant information. They then explore the possible solution space for this problem, generating as many ideas as possible, ideally without winnowing any of them out at this stage (Howard, Culley, and Dekoninck 2008). Once a critical mass has been reached, the group must make a decision. Using their judgement,

they must identify which of the ideas generated is best. This will require both evaluation and negotiation of different values and perceptions. Finally, once an idea has been decided, it must be enacted.

This holistic approach to the creative process argues that creativity is more than just the generation of ideas, but also the ability to choose the best ideas from the pool of possibility and carry it out. This is how one arrives at a creative final product. A failure at any point in this process is a failure to produce a creative output.

3.3   Explaining How Diversity Interacts with the Creative Process

It is through the creative process that researchers have primarily focused their efforts on identifying causal mechanisms between diversity and the outcomes of group collaborations. In this section, I will first identify some of the experiments from social psychology that attempt to identify these mechanisms. I will then explore the body of literature from organizational theory, engineering, and economics.

Phillips (2014) summarizes some of the recent work in social-psychological experiments on diversity and creativity to highlight the underlying mechanisms.

In a 2006 experiment, two different kinds of three-person groups were given a murder-mystery task. Some groups were made of all white students, while others were two white students and a nonwhite student. All group members shared a common for of information, while individual members were given important clues that only he or she knew. All of their combined information was needed to solve the mystery. Thus sharing information during the discussion was key to solving the murder. Groups with racial diversity performed

significantly better than homogenous groups. Phillips (2014) argues that this is because being with people similar to ourselves leads us to believe we all hold the same information and share the same perspective, slowing down the creative process.

In a 2013 experiment, participants were asked to identify themselves as either Republican or Democrat and then read a murder mystery and decide who they believed had committed the crime. Afterwards, participants were asked to write an essay communicating their perspective to another group member who disagreed with their opinion, but with whom they would need to come to an agreement. While everyone was told to convince their partner of their position, half were told that their partner was of the opposing political party and the other half were told that their partner was of the same party. Participants who were told that they had to convince a member of the same party prepared less well than participants who were told to convince a member of the opposite party. According to Phillips (2014), this is because disagreement from someone different than us prompts us to work harder. When someone thinks that another person does not share the same perspective, they perceive additional hurdles to consensus and work harder to build a stronger case for their own position.

Phillips (2014) argues that these mechanisms and benefits extend into scientific research, citing a study of 1.5 million papers from Thomson Reuter's Web of Science which found that papers written by ethnically diverse groups were associated with more citations and higher impact factors. Furthermore, citations and impact factor were positively associated with a greater number of author addresses, or geographical diversity, and a larger number of references, or intellectual diversity (Phillips 2014).

While Phillips paints a rosy picture of the clarity and consistency of the benefits of diversity, a wider review of the literature reveals a messier picture. The effects of diversity on

creativity have been mixed in the past. Many studies have indicated positive relationships between the two, while many others have indicated the opposite.

Dayan, Ozer, and Almazrouei (2017) identify a body of literature on the effects of team diversity on effective functioning of New Product Development teams as being entirely inconsistent and rife with contradictions. In their meta-analysis of 40 years of research, Hülsheger, Anderson, and Salgado (2009) identify a small positive relationship between job-diversity and innovation that was significant and a negative, but statistically insignificant relationship between background diversity and innovation. Conservatively, they concluded that they could not say that their hypotheses relating diversity and innovation were supported (Hülsheger, Anderson, and Salgado 2009). While this seems somewhat damning, it is not the whole picture. Dayan, Ozer, and Almazrouei (2017) actually argue that the inconsistency is due to several underlying mechanisms that often go unexamined. They argue that there are two main categories of diversity, functional and demographic, with opposite, non-linear relationships with diversity (Dayan, Ozer, and Almazrouei 2017). Thus, one should expect conflicting results if comparing studies using different forms of diversity, or incorrect model formulations. Harvey (2013) also described an inconsistent literature and makes a similar argument by identifying deep-level diversity that represents differences of perspectives, and although not explicitly, by extension she suggests a shallow-level diversity that does not lead to differences in perspectives. Harvey (2013) identifies divergent creativity and convergent creativity: in short, the processes underlying the generation and evaluation phases identified by Howard, Culley, and Dekoninck (2008) respectively. A consensus seems to be forming in the literature, indicating that diversity has both positive and negative effects on creativity depending on the stage of the creative process. In particular, diversity may help produce a wider range of ideas, thus potentially generating a better or

more suitable idea, but it may hinder the ability of groups to select an idea from the wider range, slowing down the process.

When it comes to the generation of ideas and exploring possibilities, a diversity of perspectives is beneficial. When it comes to reaching a decision from a wide array of options, diversity is detrimental. Richard (2017) claims that diversity improves the quality of group decision-making, the consideration of non-obvious alternatives, and making better judgements in novel situations. At the same time, he argues that diversity carries costs that make reaching consensus more difficult and reduces efficiency (Richard 2017). Richard (2017) found that companies that were trying to downsize found racial diversity detrimental, while companies trying to grow found it beneficial.

Similarly, Harvey (2013) concludes that when the creativity of a group is evaluated as a product of divergent thinking — the number, flexibility, and originality of ideas — diversity has a well-established beneficial relationship. When the creativity of a group is evaluated as a product of convergent thinking — a single creative output — the relationship becomes much less clear with contradictory results in the literature (Harvey 2013). This distinction is important. If one is evaluating creativity purely based on divergent thinking, one is stopping half way through the creative process and ignoring the rest. Creativity is not only producing a wide range of different ideas, but also being able to evaluate which ideas is best suited to the circumstances and then implementing it.

Dayan, Ozer, and Almazrouei (2017) found that functional diversity, a diversity of information relevant to a task, had an inverted U-shaped relationship with the creativity of products developed by groups. Conversely, demographic diversity, a diversity of social categorical variables like race, gender, and age, had the reverse: a U-shaped relationship. These relationships can be seen visually in Figure 3.1.

Box 20: Characterising network interactions between economic options

SYSTEM I

A   B

F   C

E   D

SYSTEM II

A   B

F   C

E   D

SYSTEM III

A   B

F   C

E   D

The intensity of network interaction is represented numerically, as follows:

/ 1   / 2   / 3   / 4

INTERACTION MATRICES

	A	B	C	D	E
B	1				
C	1	1			
D	1	0	1		
E	1	1	1	0	
F	1	1	1	1	0

	A	B	C	D	E
B	0				
C	0	0			
D	0	2	1		
E	1	1	0	0	
F	1	0	3	0	0

	A	B	C	D	E
B	1				
C	0	0			
D	0	0	4		
E	1	2	0	0	
F	1	2	0	0	1

Examples: $i_{BD} = 0$ $i_{BD} = 2$ $i_{CD} = 4$
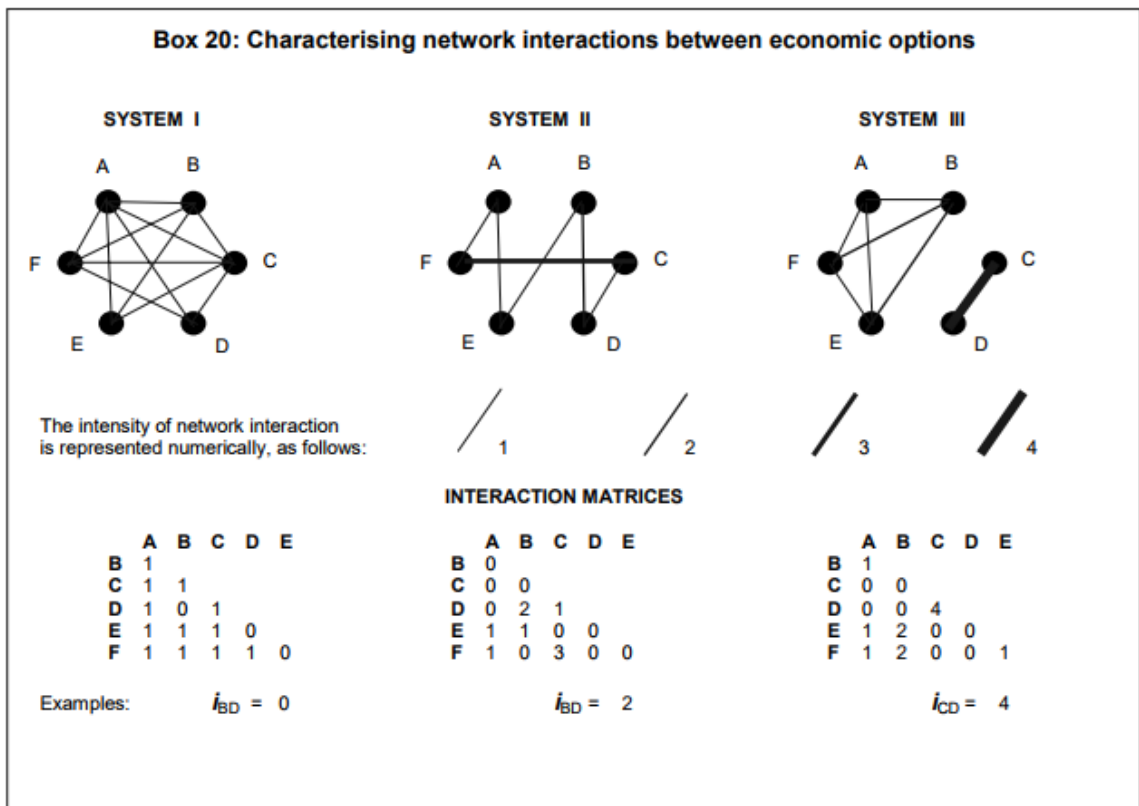
Figure 3.1: Example of Stirling's Interaction Matrices (Stirling 2006).

Where the previous authors have examined diversity by looking at the characteristics of individuals, Owen-Smith and Powell (2004) highlights another form of diversity that relates to organizations. They measure organizational diversity in Boston biotechnology groups as differences in the forms of organizations that collaborate, whether they be universities, private or public firms, but also in their missions (Owen-Smith and Powell 2004). While their definition of diversity has little bearing on this research, their ideas of the effects of diversity are particularly interesting. Owen-Smith and Powell (2004) argue that diverse networks of organizations serve as loci of innovation, where rapid innovation takes place as a result of fast learning, ease of access to information, and what they refer to as "spillover" effects of leaky collaborations. Owen-Smith and Powell (2004) examines different kinds of relationships between different actors, finding that when public research organizations anchor collaboration networks, there are spillover effects: organizations not involved in the collaborations still benefit from them because the information is visible to the whole network. This is exactly the kind of indirect collaboration that motivates part of this study.

3.4   Defining and Motivating Impact

With working definitions of diversity and creativity, and a literature that suggests several reasons that the first should affect the latter, I will explain the first group collaboration outcome that I am interested in and the means by which creativity is expected to have an effect on it: impact. I will be using impact in a neutral manner, without judging whether it is negative, positive, or a combination of the two. Any measure capable of differentiating between the two would be beyond the scope of this paper.

Garfield (2006) first suggested the idea of an impact factor in 1955 and published the

Science Citation Index in 1961. Since then, journal impact factors have been used to measure the importance, reach, or perhaps most simply, the impact of journals based on the likelihood that articles published in them will be cited. The reasoning is quite simple. If articles from a journal are cited more often, they are having a greater impact on the works that are citing them.

More recently, the Hirsch Index or h-index, has become very well known as a measure of a scientist's impact (Hirsch 2005). The h-index is given by the following definition "A scientist has index h if h of his or her N_p papers have at least h citations each and the other $N_p - h$ papers have $\leq h$ citations each" (Hirsch 2005). Unsurprisingly, an author's h-index tends to be highly correlated with the raw count of their citations (Bornmann and Daniel 2009).

While few authors are unhappy at the prospect of having more citations, and hiring committees are unlikely to look unfavourably at higher citation counts and numbers of publications, some critics have rightly pointed out that citations are not necessarily an indicator of quality or even positive attitudes. There are many reasons to cite a paper, and not all of them are complimentary. However, every measure of scientific impact invokes a count of an author's citations. Some make use of the number of papers published, like the h-index, some take into account the author's contribution to a paper based on their ranking in the list of authors, like the work of Tscharntke et al. (2007). In fact, there has been a ballooning of h-index alternatives purporting to be more accurate tools for evaluating the productivity or impact of scientists, many making explicit mention of their usefulness for evaluation committees (Anderson, Hankin, and Killworth 2008; Franceschini et al. 2012; Katsaros, Sidiropoulos, and Manolopoulos 2007; Sidiropoulos, Katsaros, and Manolopoulos 2007; Valentinuzzi, Laciar, and Atrio 2007).

For all of nuance provided by more sophisticated measures, the primary motivations are well-captured by the h-index. There is a strong sentiment that the number of citations that an author receives is a *good enough* measure of their impact, if not necessarily a perfect one. Several studies have taken similar stances when examining predictors of author and paper impacts. Some of these studies have also examined the role of diversity as one of those predictors. I will highlight some of them.

I have previously mentioned the work by Uzzi et al. (2013), linking atypical combinations of citations from different journals with a greatly increased likelihood of being one of the most highly cited papers in a field of study. Campbell and Mínguez-Vera (2008) conducted a study examining the relationship between mixed-gender coauthorship of papers and their citation rates. The authors review several articles already published that examined this issue and conclude that the literature is currently inconclusive, with different papers indicating that women are cited more often, that single male authors are cited more often, that gender has no effect, and that works with women are cited less frequently (Campbell and Mínguez-Vera 2008). However, Campbell and Mínguez-Vera (2008) hypothesizes that the mixed results in earlier studies resulted from the lack of parity in leadership roles and experience between men and women. They argued that since parity has recently been reached in ecology studies, results from the field will be better representative of the effects of gender, controlling for h-index (Campbell and Mínguez-Vera 2008).

It is important to note here that the authors argue that diversity leads to better quality work, or at least, work that peers identify as being of higher quality (Campbell et al. 2013). Impact is being measured as the number of citations received, which serves as a proxy for the community's perception of quality (Campbell et al. 2013). In this case, it is not *intellectual* diversity that produces more creative work, which then receives higher citations

as a result. Rather, the study design looks at work groups within a single discipline, and if we assume that the distribution of intellectual capital is randomly distributed between men and women in ecological studies, then we would expect it not to have an effect on the results of the study. The salient feature is *gender* diversity.

Leahey and Moody (2014) offer a different interpretation of number of citations, suggesting that they measure the usefulness of a paper because they have, in some way, had an influence on the citing paper. While I am willing to accept their assessment that number of citations is a proxy for the community's assessment of quality, for the purposes of this paper I will avoid the issue by focusing on the more defendable position that the number of citations is a proxy for the impact of a particular work.

While Leahey and Moody (2014) argue for intellectual diversity as a driving force for creativity and quality, they also test an alternative mechanism whereby intellectual diversity expands one's audience. In this way, intellectual diversity might increase the impact of a paper, not by improving its quality, but by reaching people who might not otherwise have looked at it (Leahey and Moody 2014). They conclude that integrative research does gain more citations than non-equivalent research, and the effect is not only due to having a broader appeal, indicating that both mechanisms are at play, at least in sociology, the discipline that they studied (Leahey and Moody 2014).

3.5   Defining and Motivating Productivity

While number of citations is one side of the coin that makes up the h-index, the other is the number of publications that an author has produced. There is much discussion of the current state of academia, and many references are made to the idea of "publish or perish"

(Linton, Tierney, and Walsh 2011; Qiu 2010). Researchers are incentivized to publish papers quickly and frequently, and as such, they have an active stake in understanding effects that may improve or hinder their productivity in much the same way as impact. Recalling the stages of the creative process mentioned earlier, and the hypothesized mechanism by which diversity hinders the ability of groups to quickly evaluate a wide range of ideas, productivity is thus the second outcome important to my analysis. Productivity is perhaps less contentious than impact in concept, but its measurement suffers a similar issue. Quantifying productivity has led to a wide range of different measures that have varying levels of usage in different fields (Soyer 2016). Many of these are composite measures that take into account citations received, such as the h-index, g-index, total impact factor of all publications, weighted total impact factor, number of citations, highest number of citations for a single publication, and more (Soyer 2016). I will use a much simpler measure than the ones described: productivity is the number of papers an author has published divided by the number of years they have been publishing. Thus, productivity is a measure of an author's average rate of publication. In this section, I will highlight some of the previous ideas relating diversity, creativity, and productivity.

Collins (1998) has much to say on the topic of productivity, creativity, and networks. He argues that intellectual endeavours are primarily done alone for long periods of time, but the "emotional energy" that drives those long periods of intense concentration derives from "interaction rituals" — gatherings of two or more people with a shared mood or emotion focused on a single object (Collins 1998). The shared focus and mood bring the participants into a "shared reality" that creates the experience of boundary between the situation and the outside world; this produces feelings of shared group membership and resulting obligations to one another and the group (Collins 1998). Interaction rituals charge participants with

emotional energy in proportion to the intensity of the interaction (Collins 1998).

Emotional energy is more specifically "the flow of enthusiasm that allows individuals in the throes of ritual participation to carry out heroic acts of fervor or self-sacrifice" and "charges up individuals like an electric battery, giving them the corresponding degree of enthusiasm toward ritually created symbolic goals when they are out of the presence of the group" (Collins 1998). What Collins is trying to say is that when people with a shared identity and goals gather, they build enthusiasm off one another that stays with them after the gathering ends. This enthusiasm then lets them perform more work than they would be able to otherwise. Examples of interaction rituals might include talking with a colleague over lunch — a low intensity interaction ritual — or a major conference — a high intensity interaction ritual. Depending on the individual and the intensity, differing amounts of emotional energy will be produced and consumed, thus requiring regular interaction rituals to maintain one's enthusiasm (Collins 1998). This is the mechanism that Collins argues allows social factors to influence individual productivity.

For Collins (1998), speed of publication is of utmost importance for productivity, and network position is vital to speed. He argues that cultural capital and emotional energy flows through networks and collects in individuals and groups that are ideally situated to make use of them (Collins 1998). By having access to new ideas before anyone else, and having the emotional energy to convert them into publications before anyone else, individuals can acquire an advantage over other members of the field (Collins 1998). From this, one sees an argument that a diversity of ideas and a central network position should result in faster publication rates.

Other researchers have looked into the effects of social capital on productivity and performance (Groboljsek et al. 2014). Ziherl et al. (2006) found that researchers in Slovenia

exhibited three kinds of social capital: weak social capital, where groups are small and co-operation ties are weak, strong bonding social capital, where groups are small, located in dense network structures, and co-operation ties are strong, and strong bridging social capital, where groups are larger, spread across diverse institutions, and co-operation ties are of moderate strength. They then assessed the effects of these different kinds of social capital on the performance of PhD students (Ziherl, Iglic, and Ferligoj 2006). They operationalized performance as the sum of articles, proceedings, book chapters, books, internal research reports, and participations in major conferences and workshops, with each weighted according to importance (Ziherl, Iglic, & Ferligoj, 2006). They found that the average index of PhD student performance did not vary by discipline or gender, but varied greatly by type of social capital; the average performance index was 6.63 for students with weak social capital, 9.29 for bonding social capital, and 22.13 for bridging social capital (Ziherl, Iglic, & Ferligoj, 2006). The differences between weak social capital and bonding social capital were non-significant, but the differences between both and bridging social capital were significant (Ziherl, Iglic, & Ferligoj, 2006). This indicated a non-linear relationship between the strength of co-operation ties. Group cohesion, and performance: the greatest benefit is found at moderate levels of tie-strength and group cohesion (Ziherl, Iglic, & Ferligoj, 2006). Ziherl et al. (2006) argue that the non-linear relationship between tie strength, group cohesion, and performance is characteristic of the need to balance diversity with common knowledge, a refrain that echoes the tension between novelty and conformity seen earlier. Once again, one sees an argument for diversity and network position having a positive effect on productivity.

4

Previous Measures and Studies of Diversity

In the previous section, I highlighted much of the theory, ideas, and hypotheses behind defining the important themes in this paper and the causal mechanisms that may link them together. I have described several ways that different authors have attempted to conceptualize diversity because it is directly related to their efforts to measure it. In this section, I will summarize the extensive work done by Stirling (2006) to identify three characteristics of diversity that a measure can capture and four criteria for a good measure. I will highlight some of his insights from the studies of diversity in ecology and supplement it by extending a similar analysis to several sociological measures of diversity. I will follow this by summarizing Stirling's proposed solution, an "integrated multi-criteria diversity index M and an elaboration M'. In doing so, I will highlight how these measures do not meet the needs of my study, thus justifying my introduction of two new measures for another purpose.

Stirling (2006) provides a very comprehensive overview of various measures of species diversity in ecology and their various merits in his pursuit of a good measure for economic

diversity. He highlights three characteristics of diversity that are necessary, but not individually sufficient for an generalized concept of diversity: variety, balance, and disparity (Stirling 2006).

Variety is a measure of the number of categories in which something can be partitioned (Stirling 2006). For example, the number of ethnicities that exist in a sample population. As such, variety is a positive integer. Thus, all other things held steady, a greater variety can be said to indicate a greater diversity.

Balance is a measure of the distribution of quantities in each category (Stirling 2006). For example, the number of people who identify as being of a particular ethnicity in a sample population. Balance is an array of positive fractions that sum to 1. For a given sample, the more equally distributed observations are within the categories, or the smaller the variance, the greater the diversity.

Disparity is a measure of the nature and degree to which the categories differ from each other (Stirling 2006). For example, apples to cars is a greater nature or degree of difference than apples to oranges. Disparity is an "intrinsically qualitative, subjective and context-dependent aspect of diversity" (Stirling 2006). For two samples of equal balance and variety, the one with greater disparity can be said to be more diverse.

4.1   Ecological Measures of Diversity

The majority of early measures of diversity in ecology summarized by Stirling (2006) represent measures of variety, balance, or some combination of the two. Stirling (2006) identifies 16 different measures of varying formulations which do not address disparity. In this particular category, he identifies two measures that capture the dual variety-balance nature of diversity that have received particular attention in the literature: the Simpson (or

31

Herfindahl-Hirschman concentration) and Shannon indices.

4.1.1   Simpson and Shannon Indices

The Shannon and Simpson indices are given by the respective formulas:

$$\Delta_{Shannon} = -\sum_i p_i \ln p_i$$

$$\Delta_{Simpson} = \frac{1}{\sum_i p_i^2}$$

Where p_i is the proportion of all individuals in species i.

For the above diversity indices, where the Shannon index is *higher*, the diversity of the sample is higher. Where the Simpson index is *lower*, the diversity of the sample is higher. Stirling (2006) compares the two competing indices, noting that, while some authors prefer the Simpson index because they find the exponential nature of the measure "simpler", this particular reasoning is rather subjective. Simpson's index does have two properties that would reasonably be considered detrimental to its status as a measure in comparison to Shannon's. Firstly, if one chooses a different exponential power in Simpson's index, it can produce radically different *rank orderings* for different samples; this is not the case for choosing different bases for the logarithm in Shannon's index (Stirling 2006). The second issue is more technical: in circumstances where diversity should take into account the disparity of different options of formal taxonomies, Shannon's index is a more robust basis than Simpson's because it satisfies the condition (Stirling 2006). Thus, Stirling (2006) argues that there is good reason to prefer the Shannon index over the Simpson index as a robust, generalized, and non-parametric measure of diversity encompassing variety and

balance.

4.1.2 Junge Index

In a review of diversity measures that incorporate all three properties of diversity, Stirling (2006) identifies Junge's index as a seminal, if non-robust measure, given by the following compartively daunting formula:

$$\Delta_{Junge} = \left(\frac{\sigma}{\mu\sqrt{n-1}} \right) \left(\frac{1}{\sqrt{s}} \right) \left(\sqrt{s-1} - \sqrt{s\Sigma_i p_i^2 - 1} \right)$$

Here, Junge's index is contingent on five different factors. Using Junge's language, these are: 1) the number of different classes recognized by the analyst, 2) the proportion of different cases within the classes, 3) the number of types of characteristics recognized by the analyst, 4) the standard deviation of the character values, and 5) the means of the character values (Stirling 2006). Stirling (2006) identifies many potential pitfalls here. The use of the standard deviation assumes a normal distribution, the sensitivity to minor changes in the in the values of the five factors, and the complexity of the formula itself indicates the relatively minor structural changes would result in wildly different values (Stirling 2006).

4.2 Sociological Measures of Diversity

4.2.1 Binary (Non-network)

Dezsö and Ross (2012) found that representation of women in the top management of firms was associated with increased firm performance. Using 15 years of panel data on the gender composition of the top management of the S&P 1,500 firms, they found that firms with at

least one woman in a firm's top management performed better, but only in those cases where a firm was focused on innovation as part of its strategy (Dezsö and Ross 2012). In this case, firm performance was operationalized as Tobin's Q, defined by the ratio of the market value of a firm's assets to their replacement value, chosen because it captures the firm's value holistically in a "forward-looking" manner, rather than a "backwards-looking" measure like accounting rates of return (Dezsö and Ross 2012). Since the percentage of firms with at least one woman in top management never exceeded 33%, and firms with more than one woman never exceeded 8.5%, Dezsö and Ross (2012) chose to classify gender diversity as a binary variable: 1 indicated at least one woman and 0 indicated no women in top management. Dezsö and Ross (2012) reported no significant effect of gender diversity on firm performance for firms that were not focused on innovation in either direction.

Taylor (2012) reports on analysis from the Credit Suisse Research Institute that found, out of 2,360 companies from different countries, the ones with one or more women on the board "delivered higher average returns on equity, lower gearing, better average growth and higher price/book value multiples over the last six years". Once again, gender diversity is represented as a binary variable: women are either present or they are not.

While useful for the purposes of Dezsö and Ross (2012) and Taylor (2012), binary diversity is far too simple for my purposes. It captures none of variety, balance, or disparity. Intellectual diversity is not easily characterized by binary relationships and to do so would miss out on important distinctions between degrees of similarity between the intellectual capital of authors.

4.2.2   Averaged Multi-Item Scales (Non-network)

Dayan, Ozer, and Almazrouei (2017) uses the average of a set of multi-item scales developed in earlier studies (Brown and Eisenhardt 1995; Colquitt, Noe, and Jackson 2002; Dayan and Di Benedetto 2010). Respondents were asked to rate the diversity of their team based on demographic characteristics such as age, ethnicity, and gender on a scale from 1 to 7. These responses were then averaged to provide a measure of the diversity of the team. A similar measure was used for functional, or informational, diversity using categories such as marketing, research and development, and manufacturing. These were included in a regression as both non-quadratic and quadratic interaction terms for predicting product creativity.

Dayan, Ozer, and Almazrouei (2017) were interested in testing whether informational and sociodemographic diversities operate in a similar manner on creativity. They found that while both forms of diversity had significant effects, their relationships were almost opposite. Increasing informational diversity improved creativity up to a certain point, after which increasing informational diversity was associated with decreasing creativity. Conversely, increasing sociodemographic diversity was associated with decreasing creativity until a minimum relationship was reached, after which further increases to sociodemographic diversity was associated with increasing creativity.

While Dayan, Ozer, and Almazrouei (2017)'s use of quadratic interaction terms was an important distinction from other models commonly used, their measures require surveying members of the groups. This imposes a much higher barrier for analysis and sample sizes and introduces subjective elements that may result in different evaluations of what constitutes diversity based on different interpretations of the term. Without knowing what judgements respondents are making when they interpret diversity, this measure captures

none of variety, balance, or disparity.

4.2.3   Blau's Index of Heterogeneity

This measure is in fact the same as the Simpson Index and the Gibbs-Martin Index. They differ only in whether they are measuring diversity or similarity. I include this again as a sociological example. Richard (2017) found that racial diversity was also associated with firm performance, but only to the extent that it interacted with the firm's strategies. While Richard (2017) uses racial diversity, his primary focus is using it as a proxy for cultural diversity. He hypothesizes that cultural diversity will interact with a company's strategies based on its goals (Richard 2017). He argues that cultural diversity has benefits for the quality of decision-making, consideration of non-obvious alternatives in work settings, and making better judgements in novel situations (Cox 1994; Jackson 1992; McLeod and Lobel 1992; McLeod, Lobel, and Cox 1996; Nemeth 1992). However, diversity is also associated with additional costs due to increased coordination and control (Jehn 1995; Milliken and Martins 1996; Williams and O'Reilly 1998). Blau's Index of heterogeneity bears a striking resemblance to the Simpson Index, being given by:

$$1 - \sum P_i^2$$

Here, (P) is the proportion of group members in a category, and (i) is the number of different categories represented within the unit of analysis, which in this case was a firm. Thus, if a firm has ten people, five of whom are European and five of whom are East Asian, the firm's index would be .50 (Richard 2017). A feature of Blau's index is that as the number of different categories represented increases, the maximum Blau's index increases. In a firm with five different racial categories, it's Blau's index would be 0.8 if

its members were perfectly evenly distributed between categories, or maximally heterogeneous (Richard 2017). Richard (2017) argued that cultural diversity would interact with a company's strategies due to the benefits and costs associated with it. For companies pursuing a growth strategy, increased racial diversity was associated with a significant increase in the estimated marginal mean productivity (Richard 2017). However, racial diversity was also associated with a significant decrease in the estimated marginal mean productivity for companies pursuing a downsizing strategy (Richard 2017). Richard (2017) argues that this is because firms attempting to grow into new markets are able to spend the resources necessary for diversity and in a position to make better use of the improved decision making in new situations and idea generation, while those companies pursuing downsizing require efficiency with limited resources. In short, when resources are plentiful and new opportunities are to be found, diversity is helpful for a company. When resources are scarce and the opportunities are lacking, diversity is costly.

Blau's index of heterogeneity is still too simple of a measure for my purposes. It relies on categorical variables that preclude any similarity between categories. If one were looking at the ranks of soldiers in a dining hall, Blau's index would treat each rank equally different from each other. A more nuanced approach might suggest that there is some degree of similarity between ranks that are one step removed in the hierarchy that is not seen, or seen to a much smaller degree, four or five steps removed. If one were to conceptualize intellectual diversity as the differences between the sets of books read by each author, Blau's index would be unable to account for authors reading multiple books and having partial overlap. Like the alternative variation, the Simpson Index, Blau's Index does not capture the disparity characteristic of diversity.

4.3   Stirling Index

To frame his own proposed measure of diversity that captures balance, disparity, and variety, Stirling (2006) identifies five principle elements of consistency: (i) For a portfolio of evenly balanced, equally disparate options, the diversity index should increase monotonically with variety. (ii) Where variety is equal to one, the diversity index should take a value of zero. (iii) For a portfolio of given variety and disparity, the diversity index should increase monotonically with the degree of balance in the spread of option contributions (ie: the diversity index should take a maximum value at any given level of variety and disparity where all options are represented equally). (iv) For a portfolio of given variety and balance, the diversity index should increase monotonically with the aggregate distance between options in 'disparity-space.' (v) Where this aggregate distance is zero (ie: where all options are effectively identical), the diversity index should take a value of zero.

While these criteria do not provide a sufficient basis for developing a rigorous, uniquely defined formula from first principles, they do suggest a simple algorithm that would satisfy the conditions (Stirling 2006). Stirling presents a general form of his diversity index, M, as follows:

$$M = \sum_{ij(i \neq j)} \mathbf{d}_{ij} p_i p_j$$

Where d_{ij} is some Euclidean distance metric between options i and j. Stirling (2006) summarizes this as "the integrated multicritera diversity of a portfolio of options might formally be specified as the product of the disparity-distance of each pair of options and the proportional contributions to the portfolio of each member of that pair, summed over all pairs of options." This particular formula and its properties will be important for my

formulation of group intellectual diversity.

What this means in more general terms, since Stirling is concerned with the diversity of investment portfolios, is that for given group of observations, the diversity of the group can be defined as some distance measure between two members of the group, multiplied by the proportion of the group composed of one member, multiplied by the proportion of the group composed by the other member, summed over all unique combinations of pairs in the group.

Stirling (2006) then makes an important leap from looking at cohesive groups, or cliques, to incompletely networked groups, where members of a group may interact with only some other members. Since the generalized Stirling index is already based on dyadic comparisons between group members, the leap to relational forms of data is a very short one. From this idea, he represents the interactions between the members using matrices that he terms "interaction matrices". In social network analysis, Stirling's interaction matrices would be a variant of weighted adjacency matrices where the upper right triangle is not given any values. An example of this can be seen in Figure 4.2. Note that rather than the familiar $n \times n$ form of an adjacency matrix, Stirling's interaction matrices take on $n \times n - 1$ dimensions because the rows have been shifted up, deleting the diagonal formed by self-adjacencies.

He thus extends his diversity measure, M, to a form that only accounts for such pairs where an interaction exists:

$$M' = \sum_{ij(i \neq j)} \mathbf{d}_{ij} \mathbf{i}_{ij} p_i p_j$$

Where $\mathbf{i}_{ij}$ is the corresponding value in the interaction matrix given by row i and column j. This extension of the more general Stirling index and its properties will be important
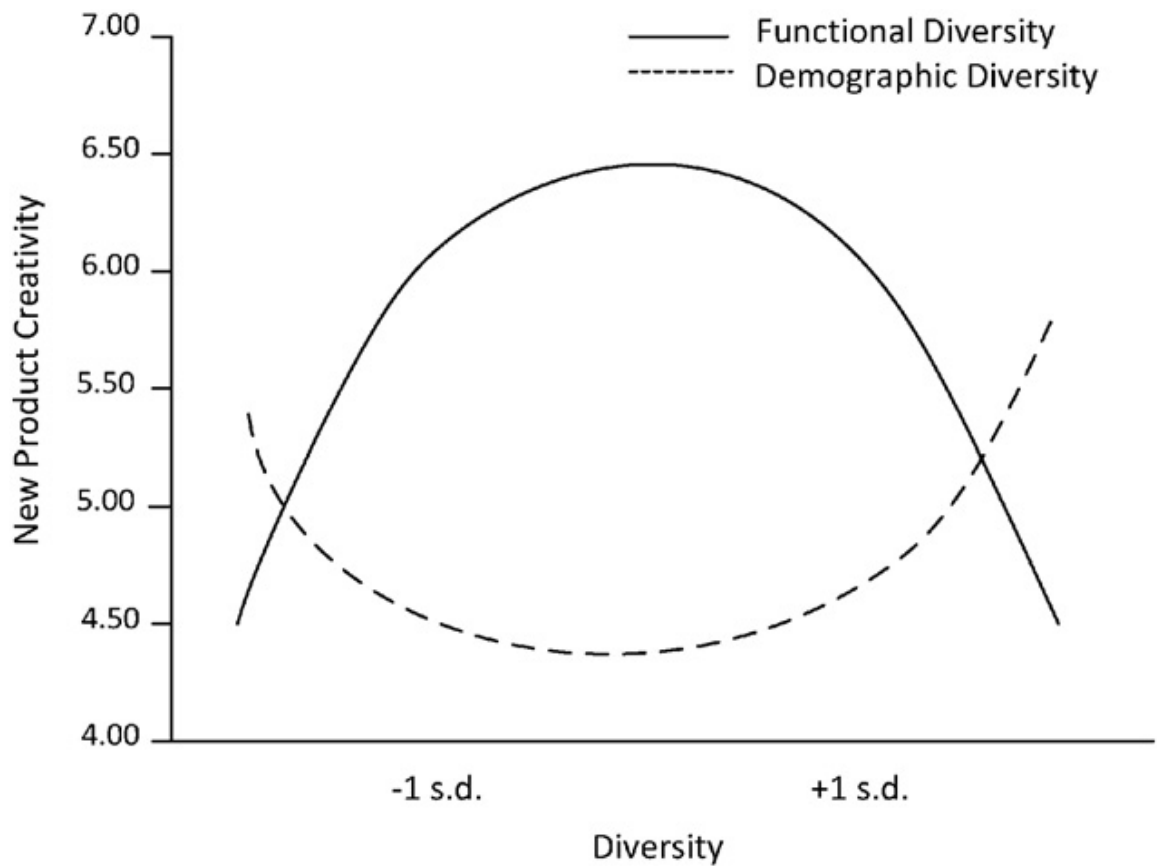
Figure 4.2: Relationship between functional and demographic diversities on new product creativity (Dayan, Ozer, and Almazrouei 2017).

for my formulation of CDS and should help understand some of the underlying ideas of network measures of diversity.

4.4   Four Criteria of a Good Measure of Generalized Diversity

Stirling (2006) argues that his index of diversity, and its extension, not only capture the three characteristics he finds important, it is superior to measures such as Junge's Index because it meets four criteria important for a good measure of diversity. He identifies these criteria as completeness, transparency, parsimony, and robustness. Stirling (2006) argues that his measure is

1) *complete* because it directly addresses the variety, balance, and disparity components of diversity.

2) *transparent* because it is non-parametric and uses a Euclidean disparity-space rather than an ultrametric geometry.

3) *parsimonious* because it only pertains to the properties that one is interested in appraising.

4) *robust* because it is not dependent on a choice of arbitrary parameter values, such as logarithm bases or exponential powers.

Stirling makes strong arguments for his measure of diversity. While the measures I propose in the next section bear strong similarities to his work, I diverge in several ways while retaining many of the strengths identified for Stirling's Index.

5

Proposed Measures

In this section, I will propose three measures. The first is a measure of intellectual capital. The second and third are measures of diversity of intellectual capital. I will provide formulas for the calculation of my diversity measures. I will then argue that they capture the characteristics of variety, balance, and disparity, as discussed by Stirling (2006). However, I account for balance differently, as an authors' presence in a cluster is binary, rather than weighted. I make a further diversion by proposing a non-Euclidean distance metric which can be normalized over a set range. I will argue that this better captures the nature of pairwise disparity between two observations and has benefits for intuitive interpretations of a group's diversity. I will conclude by showing that my measures meet Stirling's criteria for a good measure of generalized diversity.

The previous work in the field has failed to account for an individual's diversity within a group and explicitly measuring diversity in informal groups. I define the form of diversity used in this paper to be an author's intellectual capital, as measured by their *citation profile*. I propose a *coauthor diversity score* to measure how different an author is from all of their direct collaborators and *group intellectual diversity* to measure how different a group is by

summarizing the differences between each pairing of members, whether they have directly collaborated or not.

Based on the earlier works in the study of diversity summarized above, I suggest that one can expect certain relationships between diversity and the outcomes of group work. In general, one would expect: 1) increased intellectual capital would be associated with better outcomes, 2) an increase in intellectual diversity would be associated with greater impact, 3) as intellectual diversity increases, its relationship with impact should experience diminishing returns (Dayan, Ozer, and Almazrouei 2017; Howard, Culley, and Dekoninck 2008; Richard 2017). The literature on diversity and productivity is somewhat more sparse but, Ziherl, Iglic, and Ferligoj (2006) suggest a non-linear relationship between social capital and productivity, while others have suggested both positive and negative effects of diversity. From all of these expectations, I developed the following hypotheses that specifically apply to measuring diversity of citation profiles in groups of coauthors publishing in scientific fields.

5.1   Citation Profile

5.1.1   Coauthor Diversity Score

An author's coauthor diversity score (CDS) is a summary measure of how different they are from all of their coauthors. This measure provides an *individual* level metric of diversity that is often missing from other measures of diversity. Rather than capturing the diversity of an entire group, the idea behind this measure is to capture the degree of diversity that an author brings to their collaborations. Central to CDS is the idea of the difference between a pair of authors. If two authors have cited the exact same works as each other over their

careers, one would expect that they are bringing less intellectual diversity to their work than if they had cited entirely different works. In this section, I will explain how CDS is developed from pairwise diversity measures between coauthors's citation profiles using the Jaccard index.

It is helpful to think of citation profiles as sets from set theory, in that no elements are repeated. No matter how many times an author cites the same paper, it will only show up once in their citation profile. This similarity means that we can borrow the Jaccard Similarity Index from set theory to calculate the similarity of the two citation profiles. The Jaccard Similarity Index of two sets, A and B, $J(A, B)$ is given by:

$$J(A, B) = \frac{\mid Intersection(A, B) \mid}{\mid Union(A, B) \mid} = \frac{\mid A \cap B \mid}{\mid A \cup B \mid}$$

The Jaccard Similarity Index has some rather nice properties. It is normalized to range from 0 to 1, where 0 indicates the two sets share no elements and 1 indicates the two sets are exactly the same. Since I am interested in diversity, I actually need the difference between the two sets, rather than the similarity. For this, we calculate the Jaccard Distance, $d_J(A, B)$:

$$d_J(A, B) = 1 - J(A, B)$$

The resulting Jaccard Distance shares the same properties as the Jaccard Index, except reversed. A distance of 0 indicates that the two sets are exactly the same, while a distance of 1 indicates that the two share no common elements.

If one wants to determine how different someone is from all of their coauthors, one could simply take the average of their Jaccard Distances to each of their coauthors. Thus,

for an author, (i), who has worked with coauthors 1 to (n), their Coauthor Diversity Score is given by:

$$CDS_i = \frac{1}{n} \sum_{j=1}^{n} d_J(i,j)$$

Note that in this formulation $i \neq j$, which is to say that author i is never compared to themselves. This is because it makes no sense to argue that an author coauthors a paper with themselves. Even if one did, their citation profile would always be the same. Thus, a Jaccard Distance of 0 would be added and the size of n would be inflated by 1. For these reasons, I do not include an author's Jaccard Distance to themself in the calculation of CDS.

By getting an average of Jaccard Distances, which range from 0 to 1, CDS will also range from 0 to 1, providing a normalized score that allows for comparisons between authors in different groups.

5.2   Group Intellectual Diversity

While CDS is an individual-level measure that captures direct collaboration, group intellectual diversity (GID) is a summary measure of how different each member of a group is from every other member, *irrespective of whether they have directly collaborated or not*. The basic idea behind GID is very similar to that of CDS, but applied at the group level. One finds the mean intellectual difference between all possible pairings of authors in a group. It would be foolish to suggest that authors are blind to the work of their colleagues and contemporaries, even if they are not working together, thus GID does not require coauthorship ties for its calculation. This distinction will be important when examining the use

of both measures together.

For a group, A of n authors, with $d_J(i,j)$ representing the Jaccard Distance between two authors, i and j, where $i \neq j, 1 \leq i \leq n, 1 \leq j \leq n$, the *Group Intellectual Diversity* can be given by the following:

$$GID_A = \frac{\sum\limits_{i,j=1}^{n} d_J(i,j)}{\frac{n(n-1)}{2}}$$

Like CDS, GID is normalized for the number of pairwise comparisons, resulting in a range of 0 to 1 and allowing for comparisons between groups of different sizes.

5.3   Applying Stirling's Characteristics and Criteria of Diversity

Recall the three characteristics of diversity that one may attempt to capture: variety, balance, and disparity. For Stirling, who was interested in diversity of investment portfolios, it made sense for stock options to be weighted by the proportion that they contribute to their portfolio, since one may invest different proportions of one's portfolio in different options. For example, one may invest 22% of their money in oil, 55% in wind, and the rest in nuclear.

Coauthor clusters, on the other hand, are made up of binary states. An author is either in the group, or they are not. Thus, each author will always make up the exact same proportion of the cluster as any other. Since each author has equal weighting, I do not need to weight my distance metric by the proportional contributions of each author It is effectively already built into the equation.

This can be seen by comparing the generalized Stirling Index from Equation 5 and GID from Equation 10. If one weights author distances by their proportion of all distances in a group and specify a Jaccard Distance, GID is equivalent to the generalized Stirling Index.

However, the use of the non-Euclidean Jaccard Distance, rather than a Euclidean distance metric as suggested by Stirling, has an additional benefit not available in the Stirling Index. Euclidean distances have no upper bound and a lower bound of 0. The Jaccard Distance has an upper bound of 1 and a lower bound of 0. The reason for this disparity is that, for a Euclidean distance: given two sets, A and B, that share nothing in common, their distance is greater than the distance between two sets, A and C, if C is smaller than B. That is to say that two sets are more distant as the number of non-shared items increases. While this makes some sense, I argue it is less intuitive than the alternative: there is a maximum distance between two sets when they share no elements in common. In plainer English, if a person shares nothing in common with another person, they cannot be more different.

It would be a fair critique to say that one measure being more intuitive than another is not reason enough to say it is better, since intuition is subjective and it may be that for some readers, Euclidean distance is more intuitive. However, having a lower and upper bound on the difference between two sets has an additional benefit. Comparisons between distances are easier, since they are now all measured on the same scale.

Although I have focused on GID in this section, if one looks at CDS in Equation 9, one can see that the formulation is almost identical. It differs only in that for CDS, the diversity of an individual, rather than a group is being measured. As a consequence, the normalizing denominator is n rather than $\frac{n(n-1)}{2}$.

Having shown that my measures capture the three characteristics of diversity that Stirling identifies, I argue that my measures also meet his criteria for good measures of generalized diversity.

1) My measures are *complete* because they directly address the variety, balance, and disparity components of diversity.

2) My measures are *transparent* because they are non-parametric and uses more intuitive non-Euclidean distance, rather than a Euclidean distance or ultrametric geometry.

3) My measures are *parsimonious* because they only pertain to the properties that I am interested in appraising.

4) My measures are *robust* because they are not dependent on a choice of arbitrary parameter values, such as logarithm bases or exponential powers.

6

Hypotheses

Based on the earlier studies of diversity and incorporating the measures for intellectual capital and diversity proposed above, I suggest that one can expect certain relationships between diversity and the outcomes of group work. In general, one would expect: 1) an increased access to resources would be associated with better outcomes, 2) an increase in functional diversity would be associated with better outcomes, 3) as functional diversity increases, its relationship with outcomes should experience diminishing returns (Dayan, Ozer, and Almazrouei 2017; Howard, Culley, and Dekoninck 2008; Richard 2017). The literature on functional diversity and productivity is somewhat more scarse and does not lend itself to direct hypotheses. However, Ziherl, Iglic, and Ferligoj (2006) suggest a non-linear relationship between social capital and productivity, while other others have suggested both positive and negative effects of diversity. From these expectations, I developed the following hypotheses that specifically apply to measuring diversity of citations in groups of coauthors publishing in scientific fields.

1. The size of an author's *citation profile* will be positively associated with career citations.

2. *Group Intellectual Diversity* will be positively associated with career citations.

3. The positive association between Group Intellectual Diversity and career citations will have diminishing returns as Group Intellectual Diversity increases.

4. *Coauthor Diversity Scores* will be positively associated with career citations.

5. The positive association between Group Intellectual Diversity and career citations will have diminishing returns as Group Intellectual Diversity increases.

6. *Group Intellectual Diversity* will be positively associated with productivity.

7. The positive association between Group Intellectual Diversity and productivity. will have diminishing returns as Group Intellectual Diversity increases.

8. *Coauthor Diversity Scores* will be positively associated with productivity..

9. The positive association between Group Intellectual Diversity and productivity. will have diminishing returns as Group Intellectual Diversity increases.

10. The associations posited in hypotheses 1-9 will be consistent across biomechanical modelling and nanotechnology.

7

Data and Methods

To test my hypotheses, I collected data for two different data sets — biomechanical modelling and nanotechnology — to test the generalizability of my results. I chose to collect my data sets based on keyword and topic queries to identify broad areas of active research. This allowed me to work with a thematically and disciplinarily related body of work. It is important for the authors to be working in the same field to reduce the effect of different publishing and coauthoring standards in different disciplines.

I chose biomechanical modelling because I have easy access to experts in the field to help with understanding any patterns in the data. I chose nanotechnology because previous work in interdisciplinarity and diversity has been done on nanotechnology data sets. I chose fields in the natural sciences because coauthoring papers is more prevalent compared to the humanities and social sciences. Thomson Reuter's Web of Science, the database from which I collect my data, also has a bias towards the natural sciences (Mongeon and Paul-Hus 2016). Thus, restricting my data sets to the natural sciences helps maintain the reliability of my analysis.

I collected data for two different data sets to test the generalizability of my results. I

chose to collect my data sets based on keyword and topic queries to identify broad areas of active research. This allowed me to work with a thematically and disciplinarily related body of work. It is important for the authors to be working in the same field to reduce the effect of different publishing and coauthoring standards in different disciplines. For example, unlike high-energy particle physics, which can have huge numbers of authors on a paper, as witnessed at CERN, coauthorship is extremely rare in philosophy and its associated sub-fields (Healy 2015). This distinction is important because if we define the boundaries of our data to include publications from both fields, one would be examining papers with radically different coauthorship behaviours, and thus coauthorship network structures. The modularity class algorithm used in this paper and described later uses network averages to determine the whether a node is within a group or not: this assumes that similar processes underlying the formation and structure of ties across the whole network (Blondel et al. 2008).

The raw data was collected from the Web of Science database. Individual datum points consist of records: metadata on articles curated by the staff of Web of Science. These records contain a great wealth of information about the articles, but only a fraction is relevant to this work. Specifically, each record contains the following important information for an article: 1) the full title, 2) the full list of authors, 3) the full abstract, 3) the date of publication, 4) the full list of cited references, 5) the number of citations to the article.

The resulting collection of records allowed me to generate the necessary data to perform my analysis. For each individual author, I generated their citation profile by collecting every unique citation from each of the author's papers.

From each collection of records, I used the Python package *metalnowledge* (McLevey and McIlroy-Young 2017) to generate a coauthorship network representing who coauthored

a paper — from the data set — with whom. Authors may have coauthored on papers outside my data, but those works are outside the scope of this analysis. The generated network is made up of nodes, representing authors, and edges, representing a coauthorship relationship. Thus, a hypothetical paper with four authors would create four nodes, one for each author, and six edges, one edge from each author to each other author on the paper. From this network, I can then identify groups of coauthors who are more connected to each other than they are to the rest of the network using the modularity class algorithm developed by Blondel et al. (2008) to identify community structures within the network and assign each node to one of the groups. These communities of authors are the means by which I will set boundaries for what constitutes a working group of authors. When I refer to the group intellectual diversity, I am referring to the intellectual diversity of the authors within one such group. I limited my analysis to the largest 25 groups in biomechanics and 100 in nanotechnology. I decided on these thresholds to ensure a larger sample size for diversities within a group. For an example of the extreme, an author who does not coauthor necessarily cannot have a GID or CDS, and small groups run the risk of outliers skewing the measures. I chose these particular thresholds based on a visual inspection of the CDS values within groups, looking for the point at which the distribution appeared to settle on something consistent. Figure 7.3 shows the CDS histograms for each modularity class, showing a relatively consistent right-skew distribution.

Unfortunately, Web of Science does not have a means of uniquely identifying individual authors. This means that authors with the same name will be indifferentiable when creating the coauthorship network, as author names are the only key reliable enough to identify observations in the data. Authors may change affiliation, field of interest, and coauthors, so trying to identify unique authors by some combination of these variables is not feasible. In
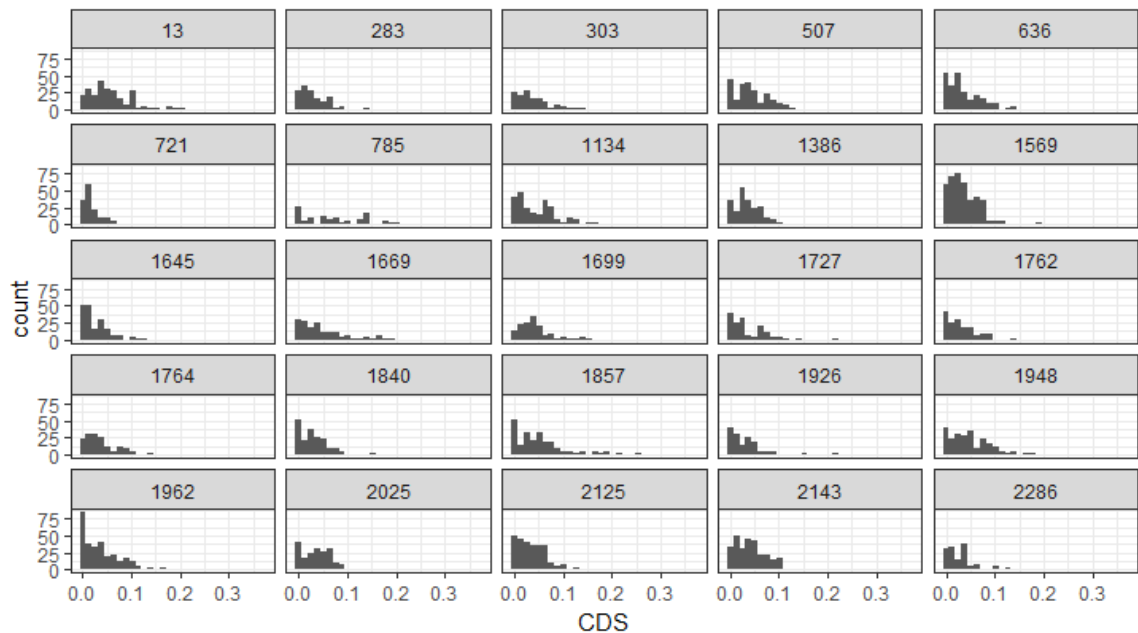
Figure 7.3: Histograms of Coauthor Diversity Scores for each Modularity Class in Biomechanical Modelling

fact, authors may also change the name under which they publish, most often by including or excluding initials for middle names or using shortened given names. To generate the network, I used the most naive method and did not parse the author names. This means that authors who spell their names differently in different publications will appear multiple times, while authors who publish under the exact same name as another author in the same field at the same time will appear to be the same person.

To reduce the amount of noise generated by author ambiguity, I have purposefully chosen smaller fields that are quite new, or else reduced the timeframe for publications that I will include in my analysis. A smaller field means a reduced likelihood of multiple authors publishing at the same time with the same name, while a smaller timeframe reduces the likelihood of past authors being confused for later authors and reduces the number of publications that an author might publish under a different name.

7.1   Biomechanical Modelling

The data set for the first case study was created from a query to Web of Science returning all journal articles containing the words "Biomechanical" and "Modelling" from the year 2006 to present in their titles, abstracts, and keywords. Since the number of publications has dramatically increased in Biomechanics, as it has in many other fields, I wanted to capture the densest period of activity while also restricting the size of the data to a manageable amount for computational purposes. I specifically chose Biomechanical Modelling because I have easy access to members of the field who have been able to provide insights into the publishing standards of the field and possible underlying mechanisms and explanations for features of the data. I restricted the date range to 2006-2017 because prior to 2006, the record availability becomes particularly sparse.

The resulting data set contained 6,543 records. From this record collection, I identified at least 21,138 authors. Each paper had, on average 4.552 authors.

From the record collection, I produced a coauthorship network with 21,138 nodes, each representing an author, and 61,374 edges, representing coauthorship relationships. For a network of size n, the network density is given by the number of edges divided by the maximum number of possible edges, which for an undirected graph is $\frac{n(n-1)}{2}$. This gives a network density of 0.0002747307. The average degree, or the average number of different coauthors that a person has in the network, is 5.807.

7.2   Nanotechnology

The data set for the second case study was created from a query to Web of Science returning all journal articles containing "nanotechnology" in their titles, abstracts, and keywords,

restricted to the Web of Science "Nanoscience Nanotechnology" category. Since nanotechnology is a relatively new field, the majority of its papers are very recent and I wanted to capture as much of the field as possible. I chose the field of nanotechnology because previous work has been done on the interdisciplinarity of nanotechnology (Porter and Youtie 2009). I restricted the date range to 2006 because prior to then, the record availability becomes particularly sparse.

The resulting data set contained 7,325 records. From this record collection, I identified at least 19,250 authors. Each paper had, on average, 5.322 authors.

From the record collection, I produced a coauthorship network with 19,250 nodes, each representing an author, and 70,160 edges, representing coauthorship relationships. Using the same calculation as above, this gives a network density of 0.0003786872. The average number of different coauthors that a person has in the network is 7.289.

7.3   Comparative Summaries of the Data Sets

When comparing two networks of different sizes, one must be very careful. For example, the combinatorial explosion of the number of possible edges in a network as the number of nodes increases makes comparing densities unwise. In general, smaller social networks tend to be more dense than larger networks formed under the same principles, for reasons that should be relatively easy to grasp.

The average degree of an undirected network is given by:

$$AverageDegree = \frac{2E}{n}$$

where, E is the number of edges and n is the number of nodes.

If one is looking at a friendship network within a class of 20 students, and students tend to have 5 friends on average (the average degree of the network is 5), then one would expect a rather dense network of 50 edges. With 20 students, the maximum number of edges is 190. Thus, we get a network density of $\frac{50}{190}$, or 0.263. If we then looked at a friendship network produced by the exact same processes, but surveyed every student in the school of 200, we would then expect a network of 500 edges. This gives us a network density of 0.025.

While this means that comparing the densities of the biomechanical modelling and nanotechnology networks is not particularly meaningful given their different sizes, it is interesting in this particular case to compare the number of edges. The nanotechnology literature has a higher rate of coauthorship, but also a higher rate of coauthoring with authors they have not coauthored with before. This distinction can be seen in the much higher average degree in nanotechnology.

Table 7.1 presents basic summary statistics for my proposed measures in the data sets for the two case studies. Authors in nanotechnology have a tendency to coauthor with authors who are more intellectually different from them, relative to biomechanical modelling. CDS values in both data sets follow a right-skew distribution. The average group intellectual diversity between the two fields is very similar, with very small standard deviations and medians close to the mean. Despite being calculated in similar manners, CDS and GID exhibit a small negative correlation in biomechanical modelling that is statistically significant. Their correlation is even smaller in nanotechnology and not significant.

	Biomechanical Modelling	Nanotechnology
CDS Mean	0.0409	0.1884
CDS Median	0.0321	0.1428
CDS Std	0.0404	0.1810
GID Mean	0.6677	0.6557
GID Median	0.6691	0.6630
GID Std	0.0141	0.0265
CDS-GID Correlation**	-0.1300***	-0.0030

Table 7.1: Comparing Summary Statistics of Measures in Biomechanical Modelling and Nanotechnology

8

Models

In this section, I provide a brief overview of the models I will be using to provide evidence for or against my hypotheses. I explain the importance and significance of including CDS and GID in the same model. Due to the dependence of observations that results from how I have collected my data and defined my variables, I violate several assumptions of my regression models. I will explain how these violations occur, why they are necessary, and what methods I will use to ensure the validity of my results despite these violations. I will then describe each model in greater detail with a summary of the variables and regression equations for ease of reference.

Recall that my hypotheses pertain to the relationships between intellectual capital, intellectual diversity, and two outcomes: impact and productivity. Thus, I need two models to assess each of these outcomes. Impact is measured using a raw count of the number of citations that an author has received from all of the papers they have been on. I will call this an author's *career citations* for simplicity. Productivity, the second outcome, is measured as the number of papers an author has published divided by the number of years between their first and last publication. I use a negative binomial regression to model career citations

59

and a linear regression to model productivity.

In both of my models, I make use of both CDS and GID. This is important because CDS in a regression will estimate the variance explained by diversity in an author's direct collaborations. GID will estimate the variance explained by the diversity of all authors to every other author in a group. Note that the definition of CDS is contained within GID. By including both terms, the variance explained by CDS is already accounted for, so GID will be explaining the diversity within groups that is *not* explained by CDS.

One can think of this as GID now estimating the relationship between an author's career citations and the pairwise diversity between every *other* author. Thus, GID in these models should be understood to represent how the diversity of a group is related to an individual's outcomes, having already accounted for the diversity the individual has brought to their collaborations.

8.1   Considerations for Statistical Analysis

Both of the models I use have an assumption that the observations are independent of one another. I violate this assumption in several ways. Because multiple authors are on the same papers, and citations are linked to papers, career citations are not independent. Furthermore, CDS and GID are both measured by making pairwise comparisons between authors in coauthoring clusters. Thus, both of those measures are reliant on the relational (non-independent) nature of networked data.

Special care must be taken to handle this violation of the independence of observations assumption. One means of doing this for linear and generalized linear models involving network data is bootstrapping the standard errors (Hanneman and Riddle 2005). Bootstrapping is an approach for evaluating the distribution of a statistic based on random sampling

with replacement (Guan 2003). Importantly, it is a non-parametric approach that makes no assumptions about the distribution of the sampled statistic (Guan 2003). The basic idea behind bootstrapping is similar to that of Monte Carlo simulations, however it resamples the original data rather than the population(Guan 2003). Provided that the sample is representative of the population, the desired statistic can be calculated as an empirical estimate of the overall population's true parameter (Guan 2003). To measure the precision of the estimates, one can calculate the bootstrapped standard errors with the following process (Guan 2003):

1. Draw multiple random samples with replacement from the sample data set.

2. Estimate the desired statistic for the bootstrap samples.

3. Calculate the sample standard deviation of the sampling distribution.

When the sample size is large, the bootstrapped estimates will converge on the population's true parameters as the number of repetitions increases (Guan 2003). Because bootstrapping makes use of a sampling of the sample, bootstrapped standard errors will be larger than those of the regular regression (Hanneman and Riddle 2005). Standard inferential formulas for computing standard errors often give unrealistically small values for network data, potentially resulting in false positives (Hanneman and Riddle 2005). Bootstrapping standard errors reduces the likelihood of rejecting the null hypothesis when one should not (Hanneman and Riddle 2005).

There is a further issue that I must resolve in my analysis. While interaction effects in linear models are generally reliable and well-understood, they are much less so for non-linear models (Drichoutis 2011; Kolasinski and Siegel 2010; Osgood, Finken, and McMorris 2002; Svensson and Oberwittler 2010). Osgood, Finken, and McMorris (2002)

61

warns that modelling interaction effects in ordinary-least-squares regressions is potentially dangerous: a mismatch between the model assumptions and skewed data can result in a greater risk of Type I and Type II errors. Svensson and Oberwittler (2010) indicates that the concerns raised by Osgood, Finken, and McMorris (2002) similarly apply to negative binomial regressions. In fact, "none of 72 economics papers on interaction effects in non-linear regression models published between 1980 and 2000 interpreted the coefficient on the interaction term correctly" (Svensson and Oberwittler 2010). Svensson and Oberwittler (2010) proposes several safeguards to ensure their results are robust and reliable:

1. Use robust standard errors.

2. Include both multiplicative and quadratic terms of the involved interaction predictors.

3. Compare the results of the regression analysis to those of Tobit regression models.

Robust standard errors help handle the heteroskedasticity of the standard errors for highly skewed data. Using the quadratic terms of interaction predictors helps account for the non-linearity of relationships. Finally, Tobit regressions have been suggested as suitable models because they can be formulated to account for left-censored data, such as would be expected in negative binomial regressions (Osgood, Finken, and McMorris 2002). If interaction effects are spurious due to non-linearity of relationships resulting from the floor effect — career citations cannot be negative — then Tobit regressions would guard against Type I errors (Svensson and Oberwittler 2010).

Thus, in order to adequately protect my analysis from the pitfalls identified above, I use a bootstrapped negative binomial regression with robust standard errors for my primary model and compare the results against a bootstrapped Tobit regression to check for Type

I errors. Since my interaction terms are already in quadratic form, I am already following safeguard 2 as proposed by Svensson and Oberwittler (2010).

8.2   Impact Model

Career citations follow a Poisson-like distribution, so I tested whether a Poisson or negative binomial regression would be more appropriate for both data sets. For the biomechanical modelling data set, career citations had a mean of 21.86 and a standard deviation of 46.06. For the nanotechnology data set, career citations had a mean of 42.65, and a variance of 92.10. In both cases, the overdispersion of the dependent variable strongly suggested that a negative binomial regression would be more suitable than a poisson regression. Furthermore, tests of the alpha parameter in the negative binomial regressions to follow showed a statistically significant non-zero value, further reinforcing the suitability of a negative binomial regression for modelling this data.

To account for the dependence of observations and issues with interaction terms in non-linear regression models, I bootstrapped the standard errors with 10,000 resamplings and included robust standard errors. As suggested by Svensson and Oberwittler (2010), I validated the model using a bootstrapped left-censord Tobit model to test whether the flooring effect of the negative binomial equation was resulting in false positives.

I included the following variables in my model:

Number of Papers - This variable is a count of the number of papers within the data set that an author has been on. It is not surprising that authors who have published more papers will tend to have more citations than authors who have published fewer papers, so this is important to control for.

Career Start - This variable is the number of years since the author's first publication

in the data set. Older papers tend to have more citations than new papers as they have had more time to accumulate citations. Simply put, a paper is more likely to have more papers five years after it was published than one year after it was published. Thus, this variable provides a means of controlling for the age of an author's papers.

Number of Coauthors - This variable is a count of the unique coauthors that have been on all papers with each author. Previous work by Uzzi et al. (2013) showed that papers coauthored by multiple authors are more likely to have higher numbers of citations. It also seems reasonable to expect that an author's citation profile size might be influenced by the coauthors on their papers, so I wanted to isolate the effects of having multiple authors and having a larger citation profile.

Citation Profile Size - This variable is a count of the unique references from all bibliographies in all papers by an author in the data set. I am using this as a proxy for the intellectual capital that an author can draw upon.

GID - This variable is the mean of the citation profile diversity scores between each combination of authors in a coauthorship group. I am using this to test the association between the intellectual diversity of a group with the outcomes of its individual members.

GID-squared - This variable is a squared interaction term for GID. I am using this to test whether differing levels of GID are associated with different outcomes for individual group members.

CDS - This variable is the mean of the citation profile diversity scores for an author with all of the other authors they have *directly* worked with. I am using this variable to test the associations between diversity with one's direct coauthors and career citations.

CDS-squared - This variable is a squared interaction term for CDS. I am using this to test whether differing levels of CDS are associated with different outcomes for individual

Variable	Name
X_1	Career Start
X_2	Number of Coauthors
X_3	Citation Profile Size
X_4	GID
X_5	GID-squared
X_6	CDS
X_7	CDS-squared

Table 8.2: Impact Regression Variables

group members.

Thus, the regression equation could be written as:

$$Impact = e^{\alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5^2 + \beta_6 X_6 + \beta_7 X_7^2}$$

Alternatively, this can be written as:

$$\ln Impact = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5^2 + \beta_6 X_6 + \beta_7 X_7^2$$

where Table 8.2 gives the appropriate variable name:

8.3   Productivity Model

Recall that I measure productivity as the number of papers an author published divided by the difference in years between their first and last publication. While intuitive, this poses an issue if an author's first publication is also their last since this would require dividing 1 by 0. As such, I used a subset of the data, removing all authors who only published a single

paper. While this dramatically reduced the size of both data sets, because my initial sample sizes were so large, I am confident that my sample sizes remain large enough for statistical power. However, my data violated the assumption of normality of residuals.

To account for the dependence of observations and issues non-normality of residuals, I bootstrapped the standard errors with 10,000 resamplings and included robust standard errors.

Career Start - This variable controls whether publication practices have changed over time.

Number of Coauthors - This variable is a count of the unique coauthors that an author has worked with. I am using this to test whether coauthoring with a wider number of authors is associated with greater or lesser productivity.

Citation Profile Size - This variable is a count of the unique references from all bibliographies in all papers by an author in the data set. I am using this as a proxy for the intellectual capital that an author can draw upon.

Career Length - This variable is a count of the length that an author has been publishing in the data set, given by the absolute difference in years since first and last publications. I am using this as a proxy for an author's career path to test whether authors at different stages in their careers have different levels of productivity.

Career Length-squared - This variable is a squared interaction term for career length. There was a very consistent u-shaped relationship between career length and productivity that needed to be accounted for in the model.

GID - This variable is the mean of the citation profile diversity scores between each combination of authors in a coauthorship group. This variable is a summary statistics for the diversity of authors in the same coauthorship group, whether they have worked

together *directly* or *indirectly* inclusive. I am using this to test the association between the intellectual diversity of a group with the productivity of its individual members.

GID-squared - This variable is a squared interaction term for GID. I am using this to test whether differing levels of GID are associated with the productivity of individual group members.

CDS - This variable is the mean of the citation profile diversity scores for an author with all of the other authors they have *directly* worked with. I am using this variable to test the associations between diversity with one's direct coauthors and productivity.

CDS-squared - This variable is a squared interaction term for CDS. I am using this to test whether differing levels of CDS are associated with the productivity of individual group members.

The equation for this model could thus be written as:

$$Productivity = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5^2 + \beta_6 X_6 + \beta_7 X_7^2 + \beta_8 X_8 + \beta_9 X_9^2$$

where Table 8.3 gives the appropriate variable name:

Variable	Name
X_1	Career Start
X_2	Number of Coauthors
X_3	Citation Profile Size
X_4	Career Length
X_5	Career Length-squared
X_6	GID
X_7	GID-squared
X_8	CDS
X_9	CDS-squared

Table 8.3: Productivity Regression Variables

9

Findings

9.1 Impact Findings

In this section, I present the results of my negative binomial regression of career citations. I summarize the findings and validate the interaction effects using a Tobit regression to test for potential false positive results. I then provide a short summary of the maximum values for the interaction effects.

Bootstrapping and robust standard errors do not change the interpretation of the model, which can be understood in the same way as other negative binomial regressions. There are several issues with interpreting interaction effects in non-linear ordinary least squares regressions, so I will avoid directly interpreting the coefficients here, instead using a Tobit regression to cross-validate the direction of effect, statistical significance, and general behaviour. I used the `margins` command from STATA to generate plots of the mean predicted value of career citations as CDS and GID change. I support this by plotting the derivative of the mean predicted value with respect to the covariate, a suggested means for examining the marginal effect of interaction terms in non-linear OLS regressions (Social Science

	Biomechanical Modelling	Nanotechnology
Intercept	-41.804*	-50.426***
Number of Papers	.073***	-.002
Career Start	.372***	.364***
Number of Coauthors	.030***	.010***
Citation Profile Size	.002***	.003***
CDS	1.298	3.047***
CDS-squared	-15.718**	-3.032***
GID	122.117*	155.787***
GID-squared	-89.899*	-118.384***
McFadden R-Squared	0.123	0.077
p	0.000	0.000
N	5157	7089

Table 9.4: Bootstrapped Negative Binomial Regression of Career Citations

Computing Cooperative 2014).

Table 9.4 presents the results of the bootstrapped negative binomial regressions for both the biomechanical modelling and nanotechnology data sets. I used STATA's *bootstrap* command with 10,000 repetitions and a seed of 42 to account for independence of observations. STATA's *robust* command was used based on the suggestion of Svensson and Oberwittler (2010) when including interaction effects in non-linear ordinary least squares models.

There were only two significant differences between the data sets. Surprisingly, the number of papers published by an author was associated with higher career citations in biomechanics, but it was *not* in nanotechnology. In biomechanical modelling, the non-quadratic CDS term did not reach the threshold for statistical significance.

The number of years since they started publishing, the number of unique coauthors, the size of their citation profile and the intellectual diversity of the cluster of coauthors they work with are all positively associated with career citations. Both of the squared interaction terms were negative.

70

	Biomechanical Modelling	Nanotechnology
Intercept	-1489.036***	-2261.294**
Number of Papers	10.011***	19.049***
Career Start	6.637***	12.032***
Number of Coauthors	.420**	.872***
Citation Profile Size	.034***	.105***
CDS	30.897	90.379***
CDS-squared	-367.830*	-86.055*
GID	4345.353**	6720.816***
GID-squared	-3273.565**	-5158.064***
Pseudo R-Squared	0.070	0.077
p	0.000	0.000
N	5157	7089

Table 9.5: Bootstrapped Tobit Regression of Career Citations

It is important to note that CDS and GID are include both on very different scales compared to the other variables, and regular and quadratic terms. Thus, interpreting them on their own is extremely ill-advised. The magnitudes of the CDS and GID terms are both a factor of the very small scale over which the two variables range as well as the change of slope as the two terms vary along that range.

Recalling the issues that arise from including interaction terms in non-linear models, one of the checks suggested by Svensson and Oberwittler (2010) is to compare the results of the non-linear model to those of a Tobit regression that has been properly left-censord. In keeping with the negative binomial, I left-censor the Tobit model at 0 and do not right-censor.

Table 9.5 presents the results of the bootstrapped left-censord Tobit regression models. Standard errors were bootstrapped 10,000 times with a random seed of 42. As before, the command for robust standard errors was included.

The direction and statistical significance of the model's terms are almost entirely identical to the negative binomial model. The only exception is that the relationship between number of papers and career citations in nanotechnology becomes significant. Recall that the primary purpose of comparing the results of the negative binomial and Tobit models is to protect against Type I errors in the interaction terms (Svensson and Oberwittler 2010). Their continued significance indicates that the floor-effect of the negative binomial regression is not producing false significant effects (Svensson and Oberwittler 2010).

While there is reason to be cautious when interpreting the coefficients for interaction terms in non-linear OLS regressions, as pointed out by Svensson and Oberwittler (2010), I have made sure to take care of the major sources of concern by using robust standard errors and cross-validating with a Tobit regression, which together provide evidence that the relationship highlighted in the negative binomial regression is not due to a false positive error. With that in mind, one can then plot the predicted mean of career citations against values of CDS and GID, holding other variables constant, using STATA's `margins` function. Note that the marginal effects of GID scores cover only a portion of the range. This is because GID scores were only observed with values between 0.55 and 0.74. Figures 9.4, 9.5, 9.6, and 9.7 give the marginal effect plots for CDS and GID in biomechanical modelling respectively and nanotechnology respectively. Their basic interpretations can be read as:

- In biomechanical modelling, increasing CDS is associated with no change in the likelihood of receiving citations until a value of about 0.15, after which it decreases.

- In biomechanical modelling, increasing GID is associated with an increased likelihood of receiving citations until a value of about 0.675, after which it decreases.
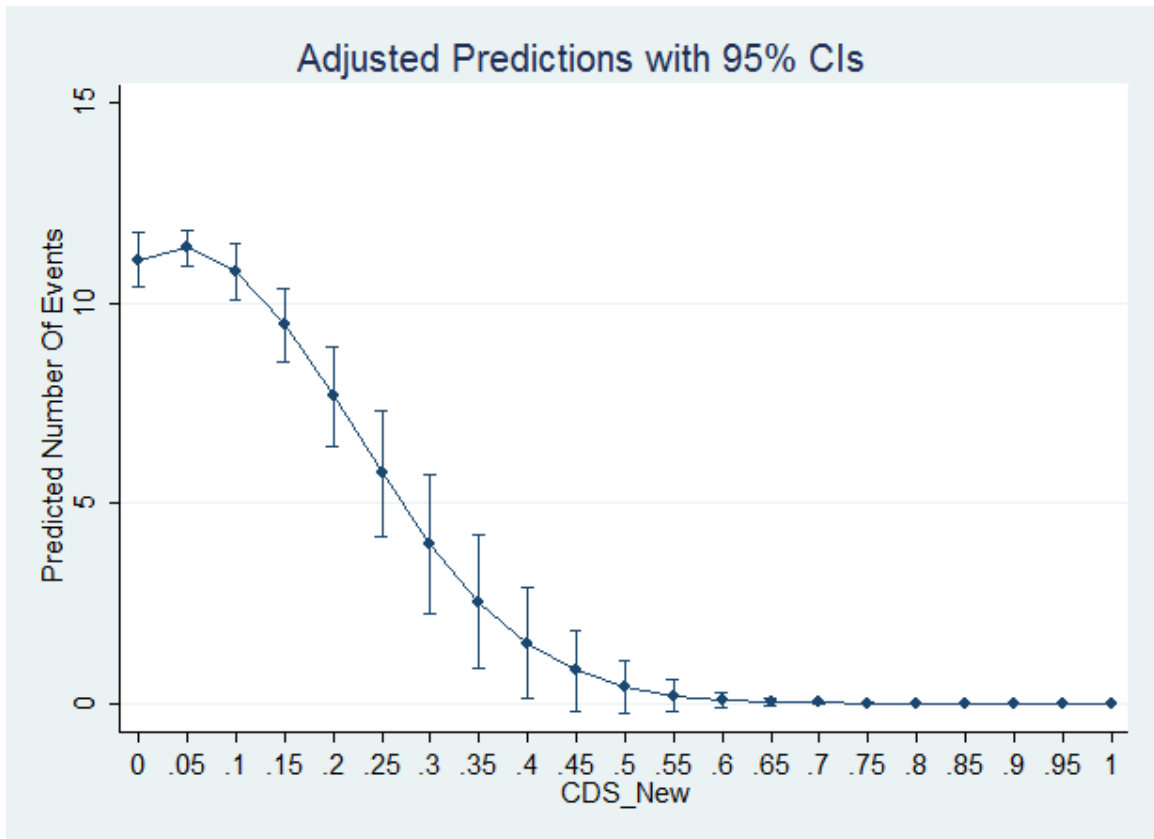
Figure 9.4: Biomechanical Modelling Negative Binomial Regression Marginal Effects: CDS

- In nanotechnology, increasing CDS is associated with an increased likelihood of receiving citations until a value of about 0.45, after which that effect decreases.

- In nanotechnology, increasing GID is associated with an increased likelihood of receiving citations until a value of about 0.65, after which that effect decreases.

Note that since the standard errors for the values around the maximum often overlap, the actual maximum may in fact be slightly different. As such, the above values should be taken as approximates. There is some concern with the biomechanical modelling CDS results, since less than 5% of observations had a CDS greater than 0.2. Thus, it may be the
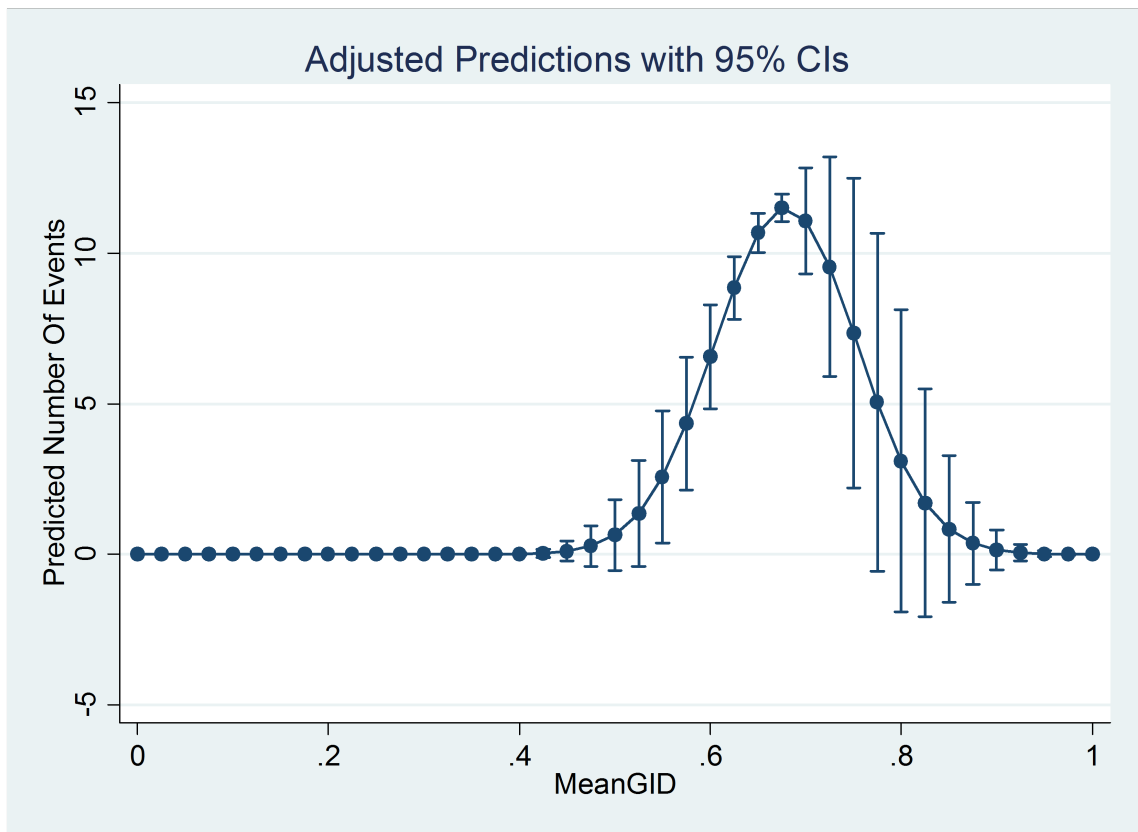
73

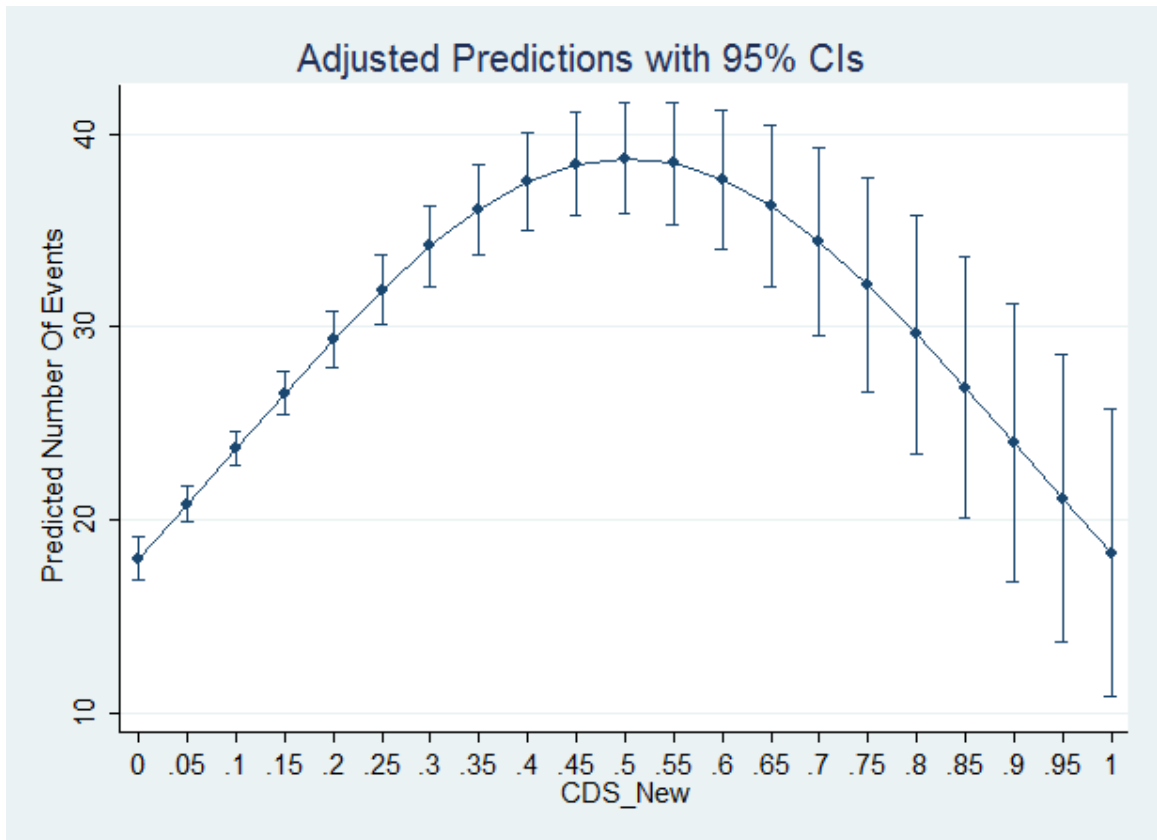Figure 9.5: Biomechanical Modelling Negative Binomial Regression Marginal Effects: GID

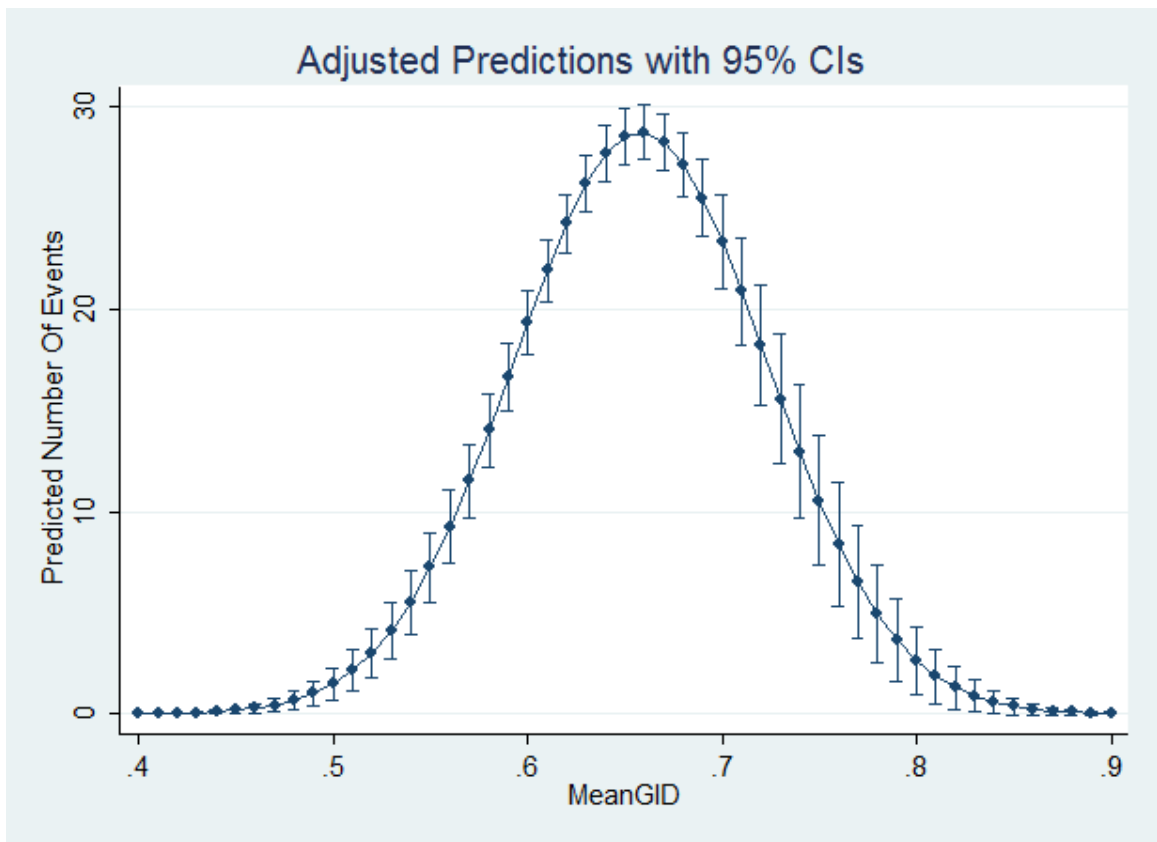Figure 9.6: Negative Binomial Regression Marginal Effects: CDS

Figure 9.7: Negative Binomial Regression Marginal Effects: GID

	Biomechanical Modelling	Nanotechnology
Intercept	-24.111**	-.062
Career Start	.012**	.006
Number of Coauthors	.022***	.016***
Citation Profile Size	.00047***	.00026***
Career Length	-.541***	-.565***
Career Length-Squared	.037***	.038***
CDS	2.792***	.365
CDS-squared	-6.605**	-.128
GID	71.879**	6.344
GID-squared	-53.230**	-5.044
R-Squared	0.663	0.734
p	0.0000	0.0000
N	1597	1454

Table 9.6: Bootstrapped Linear Regression of Papers Per Year

case that outliers are having a particularly strong effect in determining the exact slope of the relationship.

9.2 Productivity Findings

In this section, I present the results of my linear regression of papers per year. I summarize the findings and provide a short summary of how to interpret the interaction effects and the minimum and maximum values for the interaction effects.

Table 9.6 provides the results for the bootstrapped linear regression model for papers per year. Standard errors were bootstrapped 10,000 times with a random seed of 42. Robust standard errors were included.

Unlike the career citation models, the productivity models show significant differences. The year that an author first published in their respective field, their coauthor diversity

scores, and the intellectual diversity of their group of coauthors were statistically significant in biomechanical modelling, but not in nanotechnology. However, the two fields did have similar effect directions and significance levels for the number of an author's unique collaborators, the number of years that they have been publishing in the field, and its interaction term.

Interpretation of the linear model coefficients and the interaction terms poses much fewer issues than the negative binomial regressions seen earlier.

By taking the partial derivative of the regression equation with respect to one of the interacted terms, one can find the expected rate of change of productivity, holding all other variables constant. By setting this partial derivative equal to 0, one can then find the expected maximum or minimum value, after which increasing the variable's value is associated with decreasing productivity.

- In biomechanics, career length has a U-shaped association with productivity, with a minimum at approximately 0.7.

- In biomechanics, CDS has an inverted U-shaped association with productivity, with a maximum at approximately 0.2.

- In biomechanics, GID has an inverted U-shaped association with productivity, with a maximum at approximately 0.675.

- In nanotechnology, career length has a U-shaped association with productivity, with a minimum at approximately 0.75.

- In nanotechnology, CDS and GID are not significantly associated with productivity.

Note that since the standard errors for the values around the maximum often overlap,

the actual maximum may in fact be slightly different. As such, the above values should be taken as approximates.

10

Discussion

In this section, I briefly summarize the evidence for and against my hypotheses, discuss the possible meanings and interpretations of the measures used in my analysis, and provide several suggestions for what my results contribute to the research of diversity, creativity, and scientometrics.

10.1 Hypotheses Revisited

Recall that I started this analysis with the following ten hypotheses.

1. The size of an author's *citation profile* will be positively associated with career citations.

2. *Group Intellectual Diversity* will be positively associated with career citations.

3. The positive association between Group Intellectual Diversity and career citations will have diminishing returns as Group Intellectual Diversity increases.

4. *Coauthor Diversity Scores* will be positively associated with career citations.

5. The positive association between Group Intellectual Diversity and career citations will have diminishing returns as Group Intellectual Diversity increases.

6. *Group Intellectual Diversity* will be positively associated with productivity.

7. The positive association between Group Intellectual Diversity and productivity. will have diminishing returns as Group Intellectual Diversity increases.

8. *Coauthor Diversity Scores* will be positively associated with productivity..

9. The positive association between Group Intellectual Diversity and productivity. will have diminishing returns as Group Intellectual Diversity increases.

10. The associations posited in hypotheses 1-9 will be consistent across biomechanical modelling and nanotechnology.

The analysis I have performed has provided consistent evidence in support of my hypotheses, with some small exceptions. My results are mostly consistent with my hypotheses. GID and CDS both exhibited curvilinear relationships with both measures of impact. Only hypothesis (4) had mixed results, being statistically significant in nanotechnology, but not in biomechanical modelling. I provide additional evidence that these associations exhibit a maximum value after which increasing diversity becomes associated with decreasing impact. Productivity exhibits a similar relationship with my measures of intellectual diversity in the biomechanical modelling data, but exhibits no statistically significant relationship in the nanotechnology data.

10.2   Number of Papers: a measure of activity and seniority within the field

It is not at all surprising that the more papers an author publishes, the more citations they are expected to receive. By simple probability, the more papers one publishes, the more likely that one will have a paper that someone is interested in citing, all other things being equal. However, one can also view the number of papers published as a measure of activity and seniority. Many authors in the data sets had only a single paper. Whether these authors were grad students who authored a single paper with their advisor and went no further in academia, or perhaps were more established authors in other fields who dipped into the fields I looked at from a proximate field, or as a coauthor, I cannot tell. In either case, the number of papers should serve as an indication of an author's activity and seniority *within* a field.

10.3   Number of Coauthors as a Proxy for Network Position

Collins (1998) argues that authors are embedded in horizontal and vertical networks. Horizontal networks represent an author's contemporaries: rivals, acquaintances, and colleagues (Collins 1998). Vertical networks represent a genealogy of mentor-student relationships (Collins 1998). While I do not account for vertical networks, a researcher's coauthors are a major component of the horizontal networks described by (Collins 1998). Having a greater number of coauthors would be associated with a more central network position, according to Collins, conveying a wide variety of benefits (Collins 1998). For

Collins, the benefits relate to having a better view of the overall network of authors and interests, again allowing one to better capitalize on one's resources to publish new work in areas of active interest before other people do. Thus, a greater number of unique coauthors would be indicative of an author's degree centrality within a network, one of many ways that one might try to operationalize Collins' ideas. However, these are not the only benefits posited in the literature.

Uzzi et al. (2013) argues that greater numbers of authors on a paper are likely to improve the quality and reach of the paper, bringing in insights and expertise that one person alone might not have. In particular, his work focuses on the degree of novelty exhibited in papers by solo, pair, and team-authored papers (Uzzi et al. 2013).

In nearly every model, increasing the number of different people that one has coauthored was associated with positive impact or productivity. Only in the Tobit model for biomechanical modelling with it not associated with impact, but the purpose of that model was to validate the findings of the interaction terms, and should not be used to for final interpretations.

10.4   Citation Profile as a Proxy for Intellectual Capital

Collins (1998) argues that citations are an indication of the cultural capital that an author draws upon to expand the body of knowledge in which they are working. For Collins (1998), citations are both an indication of the previous knowledge being combined to create new knowledge, but also an indication of the author's knowledge of the field. One of the major arguments that Collins makes is that an author uses their cultural capital to determine where new ideas are emerging, where areas of active interest are forming, and then synthesize old knowledge into new ideas that are relevant to these areas of active interest

(Collins 1998). He stresses that everyone is trying to do this, but greater cultural capital and network positioning allows one to act faster and more effectively than others, allowing them to help define the growth of the conversation. New publications must now acknowledge their ideas since they are an early and formative part of the literature. Given that a citation profile is simply a collection of every unique cited reference from every paper by an author, it serves an acceptable role in defining one part of the cultural capital described by Collins (1998). One could thus summarize having a larger citation profile as a case of "the early bird gets the worm", where a greater knowledge of the field is like setting one's alarm clock earlier and also bringing a bigger shovel.

My analysis seems to offer some evidence to support this conceptualization of citations as indicative of cultural or intellectual capital. At the very least, authors who have more citations to different papers are more likely to have more citations over their career. While one might be concerned that this effect would be confounded by authors who write more papers, I ensured that there was no strong correlation between the size of an author's citation profile and the number of papers they had published. This allowed me to include both in my regression to try to isolate the associations between citation profile size and career citations *distinct* from the association to number of papers published.

Citation profiles might be seen as a proxy for "papers that an author has read". This may potentially be the case, but it is not necessarily so. If collaborating authors delegate sections of a paper based on expertise, one would expect that only those authors whose section of the paper makes use of the citation would have read it. Rather, I would argue that citation profiles are closer to "papers whose knowledge and ideas an author is aware of". I do not feel that it is particularly radical to suggest that authors understand the arguments within their own papers, so even if they have not read a particular paper, they likely understand

how the information in that paper is being used to support their own work.

This is key to understanding my conceptualization of citation profiles and intellectual diversity. Citations are indications of an author's awareness of how knowledge from a particular work can be leveraged to support their own ideas. While this conceptualization has the benefit of being very simple and easy to understand, it could be improved.

10.5   Coauthor Diversity Score as a Measure of the Direct Intellectual Diversity of Groups

If a citation profile is a measure of the breadth of knowledge that an author can make use of the leverage support for their own work, then an author's CDS is a measure of their tendency to work with people in their field who can leverage *different* knowledge.

The results of my analysis suggest different mechanisms, or perhaps different weightings of competing mechanisms are at play between biomechanical modelling and nanotechnology. I suggest that the benefits of negotiating different expertises may be stronger in nanotechnology, although they are still outweighed when the disparity in intellectual capital becomes too great. In biomechanical modelling, the initial flat, or slightly humped, region suggests that initially the benefits roughly match the penalties, but are eventually outweighed and increasing diversity does not improve impact.

As one becomes more different from one's coauthors, the benefit slows down and eventually reverses as the costs of translating knowledge between widely different backgrounds begin to outweigh the benefits of bringing novel information, ideas, and ways of thinking to a collaboration.

10.6 GID as a Measure of the "Leaky" Intellectual Diversity of Groups

Similar to coauthor diversity, but distinct, group intellectual diversity suggests that membership to a more diverse group, *irrespective of being oneself diverse*, has a tangible benefit. To put it succinctly, there is a "leaky" benefit to the presence of diversity within one's coauthorship group in a manner akin to that proposed by Owen-Smith and Powell (2004). Even if one is not the one introducing the diversity, nor necessarily directly collaborating with the more diverse members of the group, one still experiences the benefits. I suggest that there may be two important mechanisms that may explain the leaky benefit of intellectual diversity: the first is intangible transmission. If a sociologist were to work with a computer scientist on a sociology paper, their citation profile would likely not include many citations to computer science papers. However, they may have gained new insights, tools, and perspectives from this collaboration that they can then bring to their collaborations with other sociologists. This is a kind of second-degree effect of diversity. The second mechanism that might be at play is a visibility-due-to-proximity effect. Through one's interactions with their direct collaborators, one may be made aware of new ideas and ways of thinking that positively benefit one's own impact. Since the regression controls for coauthor diversity, GID provides an estimate of the effects of diversity *between everyone else* in the group. As such, it speaks to the indirect effects of diversity.

10.7 Productivity

The findings for the relationship between productivity and intellectual diversity are rather more mixed. In biomechanical modelling, an interesting pattern arises. While increasing

coauthor diversity is associated with some small gains, they are very marginal. Increasing gains further leads to rapidly decreasing publication rates. Conversely, productivity sees much larger gains from increasing GID until its maximum, after which increasing GID is associated with relatively minor decreases in productivity until reaching the maximum GID.

This suggests that consistently coauthoring with authors of very different intellectual backgrounds is more likely to slow down the rate of one's publishing, a conclusion well in line with the conventional argument that the more different collaborators, the harder it is to reach conclusions and solutions. Conversely, coauthoring in groups that involve a lot of diversity between the other members is primarily beneficial up to a point.

These relationships do not hold in nanotechnology, where none of these effects are significant. While the CDS and GID terms were not significant, the non-squared CDS term was almost significant. A subsequent test without the squared CDS term resulted in the linear CDS term becoming significant. This may be an indication that nanotechnology does not experience negative returns for increasing coauthor diversity. However, I am unaware of any specific differences between the activity of biomechanical modelling and nanotechnology that could explain such a difference. The insignificance in GID suggests that a group's intellectual diversity is not an important feature of nanotechnology research.

A side effect of the analysis of productivity was the discovery of the curvilinear relationship between productivity and the duration of an author's career. My analysis revealed a u-shaped distribution of productivity over an author's career in a particular field. This distribution was surprisingly consistent between biomechanical modelling and nanotechnology. It was so surprising that I compared the data to an entirely new data set collected

and cleaned in the same manner for geographical information systems literature from environmental studies and multidisciplinary geography. The exact same distribution showed up again. Given the surplus of of people with PhDs, the resulting rise of contracted positions in the academy, and one of the primary focuses of tenure being the quantity of publications, understanding what drives productivity is important not only for understanding how to make work more efficient, but also understanding how people might be trying to navigate an increasingly competitive academic job and publication market.

Such future work would do well to take into account such factors as different publication strategies, paper lengths, and publication venues such as graduate, predatory, and prestigious journals.

10.8   Takeaway

Coauthor diversity and group intellectual diversity appear to be real and consistent things that can be measured and have meaningful associations with outcomes that people care about in manners that are predicted by the literature. This paper provides empirical evidence through two cases studies for the validity of these two measures within the field of scientometrics. I contribute meaningfully to the growing literature on diversity, creativity, and group collaborations by replicating and integrating the works of theorists and previous empirical studies, while also providing a method and measure for assessing the leaky benefits of diversity, as well as empirical evidence to support their use. Across every model and every data set, there were clear and consistent benefits to coauthoring with greater numbers of different people and having access to a wider base of intellectual capital.

While it would be ill-advised to use the results of my analysis to evaluate individual researchers, I believe it offers some very general suggestions for authors. These results

should not be viewed by themselves, but rather, in conjunction with previous work that has suggested causal mechanisms between coauthorship, intellectual diversity, impact, and productivity, I suggest that my results add weight to the following recommendations:

1) Authors may find benefit from coauthoring with a wider number of researchers

2) Authors may find benefit from coauthoring with researchers who are not in their immediate area of expertise, but not authors who are excessively removed from one's research.

3) Authors may find benefit from drawing on a variety of sources to support their work. Authors should be aware of what others are publishing in their field of study, even if it may not immediately impact one's work.

11

Limitations

Because CDS and GID are generated from groups within a network, they become less meaningful when the groups are too small. The smaller the group, the more prone they are to random variation in the sample producing bias that would be smoothed out by larger sample sizes (Horváth and Kiss 2016). Similarly, authors who have only ever published by themselves are problematic for this kind of analysis. Thus, I chose for my case study disciplines that work under conditions akin to normal science, where labs are an important part of typical work and coauthoring is frequent and expected. I used a visual exploratory analysis to determine at what size the modularity classes began to assume a consistent distribution. I used the appearance of this distribution as an indication that the samples were large enough to ameliorate biases due to small sample sizes. I performed a sensitivity analysis of model results to the number of largest modularity classes by repeating my results for different thresholds. My results hold for higher thresholds and slightly lower thresholds. However, this cut-off point is ultimately based on visual inspection, rather than a rigorous analysis. I am unaware of any attempts to rigorously determine at which point sample sizes of dyadic comparisons between members of a group are large enough. Thus, my analysis

may not generalize to authors whose coauthorship clusters are smaller than my threshold.

Setting a high threshold for the number of people in coauthorship groups also ensures that the authors I am analyzing are well within the core of the field I am studying. In coauthorship-heavy fields, if a group of coauthors is too small, it may indicate a lack of engagement: either the group is relatively young, isolated, or is primarily working in a proximate field and only a few of their works show up in the data.

Since this analysis is cross-sectional in nature, I cannot make strong claims to causality in any directions. Rather than larger citation profiles leading to more career citations because they improve the work or make it more visible, it may be that impactful authors try to branch out after some initial success. It may be that authors who write impactful work tend to publish in more open fields, this openness being characterized by a greater diversity of intellectual resources being drawn upon. Impactful work might draw the attention of a wider range of authors, resulting in a subsequent boost in diversity. I suspect that these scenarios are not the case, since the bibliographies of papers definitively occur earlier in time than the citations to the papers. However, since papers are not the unit of analysis for this paper, but rather, the authors who write them, they are aggregated together irrespective of when they were written and when they received citations. Thus, I am unable to perform an appropriate time-series analysis or otherwise panelled data analysis to attempt a deeper exploration of direction and causation.

12

Future Research

The results of my analysis suggest several avenues for fruitful future exploration. While my data is currently cross-sectional, extending it to panel data would require only one additional piece of information. Determining *when* papers received each of their citations. While this is beyond the scope of this particular paper, using Web of Science's citation analysis report would provide much of the data necessary for this.

While I have focused on authors as the focus of my analysis, I believe there may be valuable insights to be gained by performing a similar kind of analysis of publications. I suspect that papers situated within topic maps may have similar relationships between how similar or dissimilar they are to other papers in their topical area.

While I have used GID and CDS together to evaluate the effects of "leaky" diversity, or diversity resulting from differences other than what one brings to the group themselves, alternative formulations of GID may be able to do this directly.

12.1   Extensions of this Paper's Ideas

I suggest that one can extend citation profiles from sets to weighted vectors. That is to say that rather than making a particular citation's presence within a citation profile a binary state of present or not, one can count the number of times that an author cites a particular paper and include that in a more nuanced weighted citation profile. This bears some resemblance to the work by Rafols and Meyer (2007), and indeed, extending the idea of the citation profile would require a different distance function that could account for the weighted nature. Cosine distance, while conceptually less parsimonious than Jaccard distance, handles this difference very well and has been used by several authors previously (Rafols and Meyer 2007, 2010; Soós and Kampis 2010).

An exciting consequence of including weights in citation profiles is that it opens up the differentiation between breadth and depth of knowledge. One can begin to differentiate between authors who have "wider" citation profiles that cover many different citations with relatively smaller weights against authors with fewer citations that they use in several papers. This opens up new avenues of studying specialist and generalist strategies in academia. Much of the research into this has been at a very high level, looking at the diffusion of knowledge across boundaries, or making arguments about core and frontier research (Kuhn 1970). By using weighted citation profiles to develop a spectrum of specialization-generalization, future research can examine individual strategies in conjunction with field and discipline-level analysis.

12.2 Generalizing the Measures to Other Forms of Diversity

While I have proposed GID and CDS as measures of intellectual diversity, as measured by diversity of citation profiles, they can be used for any kind of single or multiple-response variable. If I had gender data for the authors in my case studies, it would be trivially simple to produce a group gender diversity score or coauthor gender diversity score. The Jaccard similarity index works just as well with sets of size one. As a simple example, if one had a group of 14 men, the group gender diversity score would be 0. The average of the jaccard distance from each member to every other member is 0 because everyone is exactly alike. If one had a group of 7 men and 7 women, the group gender diversity score would be 0.462.

Note that this is not 0.5, despite there being an even number of men and women, because the GID does not compare each individual to itself. As a result, each person is compared against six people of the same gender, and 7 people of another gender. GID is the average of the difference between everyone and the other group members, not the balance of different variables. With that said, it is relatively easy to show that as the size of the sample increases, the GID of a perfectly balanced group would converge on 0.5.

Should one wish to use other distance metrics, I highly recommend using only those which can be normalized. The underlying assumptions of CDS and GID are such that there is a maximum similarity or difference.

12.3 Career Length and Productivity

As a result of an accident while formulating my model of productivity, I discovered a non-linear relationship between career length and productivity completely unrelated to my
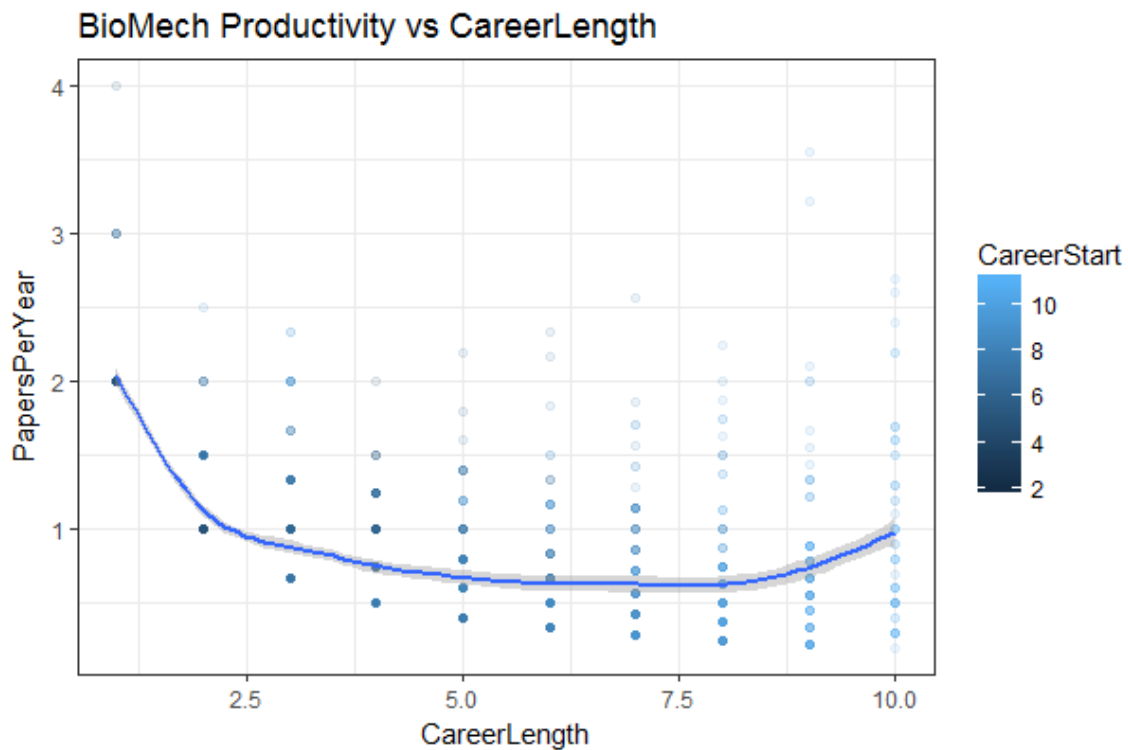
Figure 12.8: Biomechanical Modelling : Productivity vs Career Length: CDS

hypotheses. This relationship can be seen in Figure 12.8 and Figure 12.9.

This relationship was particularly surprising for two reasons. Firstly, its inclusion resulted in the greatest increase in R^2 for the model. Secondly, and more interestingly, it runs counter to the expected behaviour, as it suggests that authors are most productive right after publishing their first paper and approximately 10 years after publishing their first paper. Based on discussions with members of the field in biomechanics, this second wind, so to speak seems to be happening after the average researcher achieves tenure. Based on more traditional narratives of author productivity, one would expect productivity to increase leading up to tenure, and then decrease.

Determining the reasons behind this relationship is well beyond the scope of this paper,
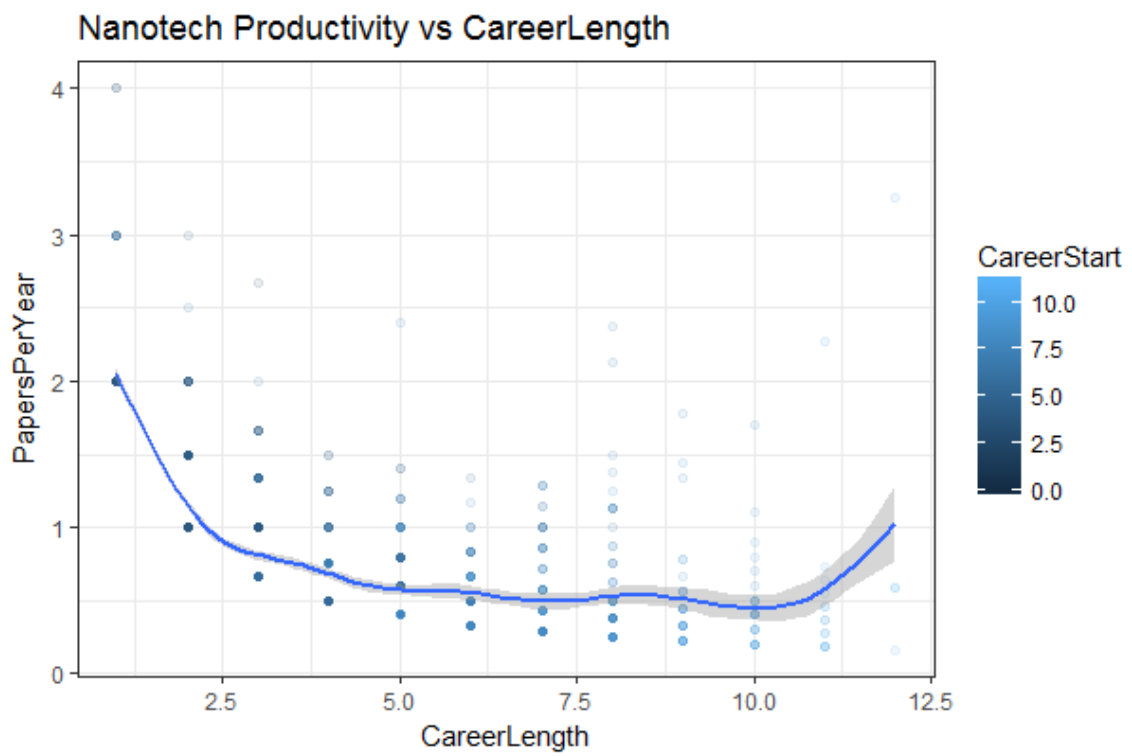
Figure 12.9: Nanotechnology: Productivity vs Career Length

but I propose two mechanisms that may be at play.

1) I suggest that the initial high productivity may be related to the practice of converting one's Masters and/or PhD theses into articles for publication. Since my measure of career length starts when an author first publishes a paper, rather than when an author starts their studies, it may be that authors are effectively front-loading a lot of the time and effort in a period before it would show up in my data. Thus, an author's appearance in the data is heralded by a burst of activity as they convert their theses into articles for publication. Once that is done, they will then fall into the normal behaviour and rate of publishing. This does not explain the spike in publications that occurs later in one's career.

2) I suggest that both high points in productivity may be associated with periods in an author's career when they have access to well-developed lab groups and resources. Since both fields are dominated by research in labs, I suggest that when an author first publishes, it is with their advisor and their lab group. This results in an increased productivity because the work is spread out between the lab, effectively boosting the rate at which individuals publish. When an author graduates and enters the academic job market, they are displaced from their original lab group and must juggle the requirements of teaching, research, and acquiring tensure. Eventually, if successful, an author will set up their own lab and be able to make efficient use of graduate students to improve productivity once again.

If this is the case, one should be able to test this by examining coauthorship patterns. One could test whether an author's publication rates are associated with well-developed and dense coauthorship clusters at the beginning of their career, which then fade for a period before a new and different cluster forms.

97

13

Conclusion

In this paper, I have introduced two measures of intellectual diversity. Coauthor diversity score (CDS) is a measure of the intellectual diversity between an author and all of their coauthors. Group intellectual diversity is a measure of the intellectual diversity of a group of authors, with the aim of capturing both direct and indirect collaborations. I have analyzed two data sets, one for biomechanical modelling and another for nanotechnology, to find associations between my measures of intellectual diversity, impact, and productivity. I present evidence that my measures are associated with statistically significant effects that one may expect based on the literature of group diversity and creativity. I show that, in many cases, increasing diversity is associated with improved impact or productivity until a maximum value is reached, after which further increases in diversity are associated with decreasing impact and productivity. The main exception to this is in nanotechnology, where productivity is not associated with my measures of intellectual diversity. Based on these results, I argue that my measures of intellectual diversity contribute to the literatures of diversity, creativity, and scientometrics by offering empirically validated, new, and generalizable measures that account for the effects of the diversity that an individual brings to

a group and the diversity of the group itself.

14

REFERENCES

Anderson, Thomas R., Robin K.S. Hankin, and Peter D. Killworth. 2008. "Beyond the Durfee square: Enhancing the h-index to score total publication output." *Scientometrics* 76(3): 577–88.

Blondel, Vincent D, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. "Fast unfolding of communities in large networks." *Journal of Statistical Mechanics: Theory and Experiment* 2008(10): P10008.

Bornmann, Lutz, and Hans-Dieter Daniel. 2009. "The state of h index research. Is the h index the ideal way to measure research performance?" *EMBO reports* 10(1): 2–6.

Brown, Shona L., and Kathleen M. Eisenhardt. 1995. "Product Development : Past Research , Present Findings , and Future Directions." *Academy of Management Review* 20(2): 343–78.

Campbell, Kevin, and Antonio Mínguez-Vera. 2008. "Gender diversity in the boardroom and firm financial performance." *Journal of Business Ethics* 83(3): 435–51.

Campbell, Lesley G., Siya Mehtani, Mary E. Dozier, and Janice Rinehart. 2013. "Gender-heterogeneous working groups produce higher quality science." *PLoS ONE* 8(10).

Castelvecchi, Davide. 2015. "Physics paper sets record with more than 5,000 authors." *Nature*.

———. 2017b. "The Large Hadron Collider."

Collins, Randall. 1998. 2 *The Sociology of Philosophy*.

Colquitt, Jason A., Raymond A. Noe, and Christine L. Jackson. 2002. "Justice in Teams: Antecedents and Consequences of procedural Justice Climate." *Personnel Psychology* 55(1): 83–109.

Cox, Taylor. 1994. *Cultural diversity in organizations: Theory, research and practice*. Berrett-Koehler Publishers.

Dayan, Mumin, and C. Anthony Di Benedetto. 2010. "The impact of structural and contextual factors on trust formation in product development teams." *Industrial Marketing Management* 39(4):

Dayan, Mumin, Said Elbanna, and Anthony Di Benedetto. 2012. "Antecedents and consequences of political behavior in new product development teams." *IEEE Transactions on Engineering Management* 59(3): 470–82.

Dayan, Mumin, Muammer Ozer, and Hanan Almazrouei. 2017. "The role of functional and demographic diversity on new product creativity and the moderating impact of project uncertainty." *Industrial Marketing Management* 61: 144–54.

Dezsö, Cristian L., and David Gaddis Ross. 2012. "Does female representation in top management improve firm performance? A panel data investigation." *Strategic Management Journal* 33(9): 1072–89.

Distrowatch. 2017. "Distrowatch.com." `https://distrowatch.com/`.

Drichoutis, Andreas. 2011. "Interpreting Interaction terms in Linear and Non-Linear Models: A Cautionary Tale." *IDEAS Working Paper Series from RePEc*: 15–25.

European Organization for Nuclear Research. 2015. *CERN Annual Report 2015*.

Foundation, The Linux. 2013. "Linux Kernel Development: How Fast it is Going, Who is Doing It, What They are Doing, and Who is Sponsoring It [2013]."

Franceschini, Fiorenzo, Maurizio Galetto, Domenico Maisano, and Luca Mastrogiacomo. 2012. "The success-index: An alternative approach to the h-index for evaluating an individual's research output." *Scientometrics* 92(3): 621–41.

Garfield, Eugene. 2006. "The History and Meaning of the Journal Impact Factor." *journal of the american medical association* 19104(1): 90–93.

Groboljsek, Blanka, Anuska Ferligoj, Franc Mali, Luka Kronegger, and Hajdeja Iglic. 2014. "The role and significance of scientific collaboration for the new emerging sciences: The case of Slovenia." *Teorija in Praksa* 51(5): 866–85.

Guan, Weihua. 2003. "From the help desk: Bootsrapped standard errors." *Stata Journal* 3(1): 71–80.

Hanneman, R.a., and M. Riddle. 2005. "Introduction to social network methods." *Network*: 149.

Harvey, Sarah. 2013. "A different perspective: The multiple effects of deep level diversity on group creativity." *Journal of Experimental Social Psychology* 49(5): 822–32.

Healy, Kieran. 2015. "Gender and citation in four general-interest philosophy journals, 1993-2013."

Hirsch, J. T. 2005. "An Index to Quantify an Individual's Scientific Research Output." 102(46): 16569–72.

Horváth, Gergely, and Hubert János Kiss. 2016. "Correlated observations, the law of small numbers and bank runs." *PLoS ONE* 11(4).

Howard, T. J., S. J. Culley, and E. Dekoninck. 2008. "Describing the creative design process by the integration of engineering design and cognitive psychology literature." *Design Studies* 29(2): 160–80.

Hülsheger, Ute R., Neil Anderson, and Jesus F. Salgado. 2009. "Team-level predictors of innovation at work: A comprehensive meta-analysis spanning three decades of research." *Journal of Applied Psychology* 94(5): 1128–45.

Jackson, Susan E. 1992. "Consequences of group composition for the interpersonal dynamics of strategic issue processing." *Advances in Strategic Management* 8: 345–82.

Jehn, Karen A. 1995. "A Multimethod Examination of the Benefits and Detriments of Intragroup Conflict." *Administrative Science Quarterly* 40(2): 256.

Katsaros, Dimitrios, Antonis Sidiropoulos, and Yiannis Manolopoulos. 2007. "Age decaying H-index for social networks of citati." In *CEUR Workshop Proceedings*,

Kern, Scott. 2011. "Analytic model for academic research productivity having factors,

interactions and implications." *Cancer Biology & Therapy* 12(11): 949–56.

Knapp, Alex. 2012. "How Much Does It Cost To Find A Higgs Boson?" *Forbes*.

Kolasinski, Adam C., and Andrew F. Siegel. 2010. "On the Economic Meaning of Inter-action Term Coefficients in Non-Linear Binary Response Regression Models." *Working paper, University of Washington*.

Kuhn, T. 1970. *The Structure of Scientific Revolutions*.

Leahey, Erin, and James Moody. 2014. "Sociological Innovation through Subfield Integra-tion." *Social Currents* 1(3): 228–56.

Linton, Jonathan D., Robert Tierney, and Steven T. Walsh. 2011. "Publish or Perish: How Are Research and Reputation Related?" *Serials Review* 37(4): 244–57.

Mannix, Elizabeth, and Margaret A. Neale. 2005. "What differences make a difference? The promise and reality of diverse teams in organizations." 6: 31–55.

McLeod, P. L., S. A. Lobel, and T. H. Cox. 1996. "Ethnic Diversity and Creativity in Small Groups." *Small Group Research* 27(2):

McLeod, Poppy Lauretta, and Sharon Alisa Lobel. 1992. "The Effects of Ethnic Diversity on Idea Generation in Small Groups." *Academy of Management Best Papers Proceed-ings* 8(1961): 227–31.

McLevey, John, and Alexander V Graham. "Network Structure and the 'Chain of Trust' in Free/Libre Open Source Software Development: The Case of Linux." *Working Paper*.

McLevey, John, and Reid McIlroy-Young. 2017. "Introducing metaknowledge: Software for computational research in information science, network analysis, and science of

science." *Journal of Informetrics*. 11:176-197.

Menzies, T., L. Williams, and T. Zimmermann. 2016. *Perspectives on data science for software engineering*.

Milliken, Frances J., and Luis L. Martins. 1996. "Searching for common threads: Understanding the multiple effects of diversity in organizational groups." *Academy of Management Review* 21(2): 402–33.

Mishra, Punya, Danah Henriksen, William Cain, Chris Fahnoe, Kristen Kereluik, Colin Terry, and Laura Terry. 2013. "A NEW Approach to Defining and Measuring Creativity: Rethinking Technology & Creativity in the 21st Century." *TechTrends* 57(5): 10–13.

Mongeon, Philippe, and Adèle Paul-Hus. 2016. "The journal coverage of Web of Science and Scopus: a comparative analysis." *Scientometrics* 106(1): 213–28.

Nemeth, CJ. 1992. "Minority dissent as a stimulant to group performance." *Group process and productivity* (January 1992): 95–111.

Newton, Isaac. 1676. "Letter from Sir Isaac Newton to Robert Hooke.": 3.

Osgood, D. Wayne, Laura L. Finken, and Barbara J. McMorris. 2002. "Analyzing multiple-item measures of crime and deviance II: Tobit regression analysis of transformed scores." *Journal of Quantitative Criminology* 18(4): 319–47.

Owen-Smith, Jason, and Walter W. Powell. 2004. "Knowledge Networks as Channels and Conduits: The Effects of Spillovers in the Boston Biotechnology Community." *Organization Science* 15(1): 5–21.

Phillips, Katherine W. 2014. "How Diversity Makes Us Smarter." *Scientific American*.

Polanyi, Michael, and Marjorie Glicksman Grene. 1969. *Knowing and Being Essays*. University of Chicago Press.

Porter, Alan L., and Jan Youtie. 2009. "How interdisciplinary is nanotechnology?" 11: 1023–41.

Qiu, Jane. 2010. "Publish or perish in China." *Nature* 463(7278): 142–43.

Rafols, Ismael, and Martin Meyer. 2007. "How cross-disciplinary is bionanotechnology? Explorations in the specialty of molecular motors." *Scientometrics* 70(3): 633–50.

———. 2010. "Diversity and network coherence as indicators of interdisciplinarity: Case studies in bionanoscience." *Scientometrics* 82(2): 263–87.

Richard, Orlando C. 2017. "Racial Diversity , Business Strategy , and Firm Performance : A Resource-Based View Author (s): Orlando C . Richard Source : The Academy of Management Journal , Vol . 43 , No . 2 (Apr ., 2000), pp . 164-177 Published by : Academy of Management Stable." 43(2): 164–77.

Sidiropoulos, Antonis, Dimitrios Katsaros, and Yannis Manolopoulos. 2007. "Generalized Hirsch h-index for disclosing latent facts in citation networks." *Scientometrics* 72(2): 253–80.

Social Science Computing Cooperative. 2014. "Exploring Regression Results using Margins."

Soós, Sándor, and George Kampis. 2010. "Diversity and Polarization of Research Performance : Evidence from Hungary." *CoRR* abs/1009.5:

Soyer, P. 2016. "Academic Productivity." *Diagnostic and Interventional Imaging*.

Stirling, Andrew. 2006. "On the Economics and Analysis of Diversity." *Science Policy Research Unit (SPRU), Electronic . . .*: 141.

Svensson, Robert, and Dietrich Oberwittler. 2010. "It's not the time they spend, it's what they do: The interaction between delinquent friends and unstructured routine activity on delinquency. Findings from two countries." *Journal of Criminal Justice* 38(5): 1006–14.

Taylor, Madeline. 2012. "The gender gap : A quota for women on the board."

The Linux Foundation. 2017. "What is Linux?"

Tscharntke, Teja, Michael E. Hochberg, Tatyana A. Rand, Vincent H. Resh, and Jochen Krauss. 2007. "Author sequence and credit for contributions in multiauthored publications." 5: 0013–0014.

Uzzi, B., S. Mukherjee, M. Stringer, and B. Jones. 2013. "Atypical Combinations and Scientific Impact." *Science* 342(6157): 468–72.

Valentinuzzi, M E, E Laciar, and J L Atrio. 2007. "Two new discipline-independent indices to quantify individual's scientific research output." *Journal of Physics Conference Series* 90(1): 012018.

Williams, Katherine Y., and Charles O'Reilly. 1998. "Demography and Diversity in Organizations: A Review of 40 Years of Research." In *Research in Organizational Behavior*, 77–140.

Ziherl, Petra, Hajdeja Iglic, and Anuska Ferligoj. 2006. "Research Groups' Social Capital: A Clustering Approach." *Metodoloski zvezki*.

Ziller, Robert C., Richard D. Behringer, and Jacqueline D. Goodchilds. 1962. "Group creativity under conditions of success or failure and variations in group stability." *Journal of Applied Psychology* 46(1): 43–49.