Università degli Studi di Torino

Vilfredo Pareto Doctoral Program in Economics

XXXIII Cycle

# Essays on the Economics of Science

by

Daniel Fernando de Souza

**Supervisor**:
Professor Magda Fontana
Department of Economics and Statistics "Cognetti de Martiis"
University of Turin

© Daniel de Souza, Turin, Italy, 2022

# Essays on the Economics of Science

by
Daniel Fernando de Souza

A thesis submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Economics

## Abstract

This thesis studies the recent shifts in the production of academic research towards more interdisciplinary and collaborative works, exploring the increasing pace of its diffusion through both the traditional means and on social media. I study this from an economics and complexity science perspective, using publication and citation data combined with novel data sources coming from university records and a social media platform. The chapters of this thesis shed light on the costs and benefits of taking a interdisciplinary research career path, collaborating in tightly knit research groups and disseminating research online.

The first chapter – co-authored with Magda Fontana, Martina Iori and Valerio Leone Sciabolazza – studies how the choice to conduct interdisciplinary work affects a researcher's career. Using data on 23,926 articles published by 6,105 researchers affiliated with the University of Florida in the period 2008-2013, we show that synthesizing knowledge from diverse fields pays off in terms of reputation. However, if combining too-distant research fields, the impact of a work is penalized. Moreover, research conducted balancing the contribution of different scientific fields has a negative impact on the reputation of scientists in terms of the number of citations but a positive impact on the diffusion of knowledge across other disciplines. Our findings are robust to a number of controls, including individual, time, and field of study fixed effects, and they apply to all investigators regardless of their gender, collaboration behavior, performance, and affiliation. All in all, despite its public benefits, interdisciplinary research comes with a cost for a researcher's academic career. This trade-off poses challenging questions to policymakers.

The second chapter – co-authored with Magda Fontanta, Martina Iori and Valerio Leone Sciabolazza – studies the interplay between different structures of research collaboration and scholars' research portfolio diversification. Using data on 2,446 researchers at the University of Florida who co-authored with a colleague in the

period 2008-2013, we investigate how the tightness of researchers' team of collaborators is related to the level of interdisciplinarity of their publications. We find that researchers who collaborate in highly clustered groups publish in a less diverse pool of disciplines. We also find evidence that an increase in the number of co-authors amplifies the effects of close-knit teams. We provide suggestive evidence that our results are driven by risk aversion. These findings imply that besides the well-documented institutional bias against interdisciplinarity, the internal dynamic of teamwork may hinder the adoption of a more diverse research agenda.

The third chapter studies how the academic-related activity of the most popular economists on Twitter is related to their research outcomes. In particular, I ask if wider research dissemination on Twitter is related to an improvement in scholars' citations and publication metrics. In order to answer these questions I construct a novel dataset of social media activity comprising of the 471 most followed economists' profiles on Twitter who have an academic publication record to answer these questions. I find that sharing scientific papers on Twitter is associated with an increase in individual citations. I also find that despite female scholars being on average more influential in the economists' Twitter network, they are less likely to disseminate research on Twitter and take advantage of their network position to further their academic careers.

# Acknowledgements

My PhD journey was a long and tortuous one – as it is often the case for research aspirants like me all around the world. In my view, one of the most beautiful aspects of pursuing a doctoral degree is the sense of shared experience, realizing that some of the challenges you face are similar to those dealt by other doctoral students in different fields, universities, countries and even across generations. This realization gave me the sense of being part of something bigger than myself, a feeling that constantly motivated and fulfilled me over the last four years. Despite the exhaustive and nerve-racking daily routine I endured to write this thesis, I feel extraordinarily privileged to be able to pursue a scientific endeavour, to be a member of centenary institutions and to be able to explore and satisfy this timeless and very human desire called curiosity. With that being said, what really made my journey special was the people who walked this path by my side. It is my pleasure to thank those who helped and supported me along the way.

A owe a deep sense of gratitude to my supervisor Magda Fontana for her compassionate, encouraging and thoughtful guidance. I would not be able to finish this work without her support and advice in the academic and personal level. Being her advisee and co-author is one of the academic honors I will always carry with me. I would also like to thank my co-author Valerio Leone Sciabolazza, for his patience, kindness, meticulous scrutiny and scholarly advice which deeply inspired my research work. Lastly, I am deeply grateful to my co-author Martina Iori for pushing me forward, never giving up, offering me good advice and for the genuine friendship we develop over the years – I would not have survived through these daunting years of graduate school without it.

I would also like to thank the many current and former colleagues at the University of Turin that I have had the pleasure to meet during this period. In particular, Beatrice Braut and Angelo Petralia were fundamental the first year of courses in the Collegio Carlo Alberto. Their presence brightened these stressful days of relentless study and their openness and lasting friendship were (and still are) essential for my adaptation to Italian life and culture. I am also grateful to Fabrizio Fusillo, Francesco Trentini, Anna Velyka and Carlo Bottai for the warm welcome and companionship in the sharing office at Campus Luigi Einaudi. Thanks also to Judy Song, Altayesh Terefe and Zviko Chadambuka for sharing the joys and frustrations of being a foreign PhD candidate in Italy. I am also extremely grateful to Stefania Basiglio and Giulia Briselli who started as colleagues but became the closest thing to family I had during the troubling pandemic - I am lucky to be able to call them friends. A debt of gratitude is also owed to Giulia Chersoni – I am fortunate to have worked closely and shared these years as a graduate student with her. She was a source of strength in the toughest moments, a trustworthy friend and colleague who brought joy and lightness to seeming impossible tasks and situations.

I would also like to thank my friends in Brazil, Italy and elsewhere, without whom I would not be here. There are too many friendships to list, but I would certainly be remiss not to mention a few names. First, I would like to thank Lizia de Figueirêdo, my first academic mentor, supporter and cheerleader. Lizia has been the earliest and most influential force in my academic life and the friendship we built is very precious to me. I would like to give a heartfelt thanks to my dear friends Gustavo Roscoe, Audrei Gonzaga and Fúlvio Sampaio for the life long rapport that has shaped me into the person I am today. I am also grateful by the constant presence of Angelo Zortea, who despite the distance shared all the good and bad moments of my Italian experience. I also owe a big debt of gratitude to Luigi, Mattia and Mariachiara who made my life in Piedmont so delightful. Finally, I would like to thank Bertha Rohenkohl for being such considerate and kindhearted friend. Every step forward in my PhD path was given knowing that I could count on her immeasurable support, insightful advises and boundless friendship.

My accomplishments were only possible due the support of my family. My mom Cássia Valéria is always going to be an example of resilience, strength and love. It was a privilege to be raised by a woman who thrived despite a life of hardship. I would also like to express my immense gratitude to my dear uncle Clauder for being present in my life and providing all the assistance he could from my childhood until now. Lastly, I am deeply grateful to my brother Lucas for being my main source of inspiration. His example of perseverance, dedication and selflessness has always made me strive for progress. Thanks for always believe and supporting me, you are the best brother anyone could wish for.

# Contents

This page intentionally left blank

# List of Figures

This page intentionally left blank

# List of Tables

# Chapter 1

# The Interdisciplinarity Dilemma: Public versus Private Interests

***Joint with Magda Fontana, Martina Iori and Valerio Leone Sciabolazza***

## 1.1 Introduction

In recent years, diverse patterns have emerged in science. Scientists have narrowed their expertise in response to the burden of knowledge (Jones, 2009) and rely more and more on teamwork by joining different fields of specific knowledge – interdisciplinary research (IDR) – to produce wide-ranging scientific advances (Cedrini and Fontana, 2018; Larsen and Ins, 2010). The growing importance of interdisciplinarity also results from the push of private and public funding and research institutions that find the overcoming of disciplinary barriers (Rylance, 2015) as the optimal solution to scientific and social problems. These new patterns have produced profound changes in the organization of science: universities created interdisciplinary research centers (Biancani et al., 2018; Wuchty et al., 2007; Hackett et al., 2021), and science as a whole has experienced an increasing trend of citation flows across disciplines in several fields of study (Angrist et al., 2020; Battiston et al., 2019). Moreover, studies have found that IDR is associated with more grant and patent submissions and with stable cooperation networks (Arnold et al., 2021; Jha and Welch, 2010; Singh and Fleming, 2010).

Thus, it might seem that interdisciplinarity is the optimal response to the ongoing transformation of science. On one hand, it counteracts the effects of specialization by allowing researchers to join expertise and to face more challenging societal and scientific issues and therefore fulfilling the public interest to face the complexity of societal problems that increasingly require expertise from different fields. On the other hand, the relative abundance of funding available to undertake IDR (Singh and Fleming, 2010) makes interdisciplinarity a sensible option for scholars. However, recent literature (Arnold et al., 2021; The National Academies, 2005) has raised doubts about the presence of a potential conflict between the private interest of the researcher (career and reputation) and public interest (solution of societal issue and circulation of knowledge beyond disciplinary boundaries). Actually, the reorganization of the academic community towards receiving interdisciplinary is far from being completed. Scholarships and their assessment mechanisms are still organized in separated disciplines or even in sub-fields. The specialization of journals (Stigler et al., 1995), together with the decreasing importance of generalist journals (Goel and Faria, 2007, p. 538), suggests that academic reputation tends to be built within niches. Moreover, the increasing relevance of rankings of field-specific journals renders the interdisciplinary effort rather risky since these rankings are used to evaluate research performances of universities, departments, and individual scholars and, then, to assign funds and make hiring decisions (Cedrini and Fontana, 2018; Ritzberger, 2008).

In this paper, we explore the idea that researchers often receive contrasting incentives when conducting their work. While an interdisciplinary approach is required to produce scientific advances and access to funding, the academic scholarships and evaluation mechanisms are still organized following the criteria of traditional disciplinary fields. If pursuing interdisciplinary research results in contrasting outcomes, science may face an interdisciplinarity dilemma: should researchers pursue their own private interest to build a reputation? Or should they endeavor towards public interest? How costly in terms of reputation is to choose IDR over (more) specialized research? To investigate the trade-off, we study, at the researcher level, the effect of adopting an interdisciplinary approach: i) on the number of citations received by researcher's papers, as a proxy for reputational achievement; ii) on the circulation of researcher's papers across diverse fields, as a proxy for the public interest to face societal issues through the circulation of expertise beyond disciplinary boundaries.

Namely, we aim in this paper at: i) contributing to the literature on IDR by adding to the scant evidence on the effect of IDR on the researchers' career (Leahey et al., 2017; Sun et al., 2021); ii) filling a gap in the extant analyses of the topic: previous research shows mixed evidence on how interdisciplinarity affects scientific impact – number of citations –, productivity, and research funding (Leahey et al., 2017; Sun et al., 2021) but an analysis of the trade-off between private and public interests is yet to be explored.

Toward this purpose, we analyze a novel and unique dataset of 6,105 researchers affiliated with the University of Florida (UF) along with their publication records (23,926 articles) and individual characteristics (such as gender and affiliation) over the period 2008-2013.[1] Albeit small in comparison with the samples used in other studies (Yegros-Yegros et al., 2015), our dataset has the unique feature of providing detailed bibliometric and non-bibliometric information about a panel of scholars operating in a wide range of scientific fields and affiliated to the same university. This feature allows sorting out a number of confounding factors often neglected by the literature, as the role played by institutional and national heterogeneity.

Thanks to the panel nature of data, we observe the variation of the degree of interdisciplinarity across articles by the same scholar.[2] With respect to extant literature (see, for instance, Yegros-Yegros et al., 2015), we account for the investigators' individual characteristics that may play a crucial role in determining the scholar's reputation. At the same time, by performing our analysis at the article level and comparing papers of the same researcher (through individual fixed effect), we avoid aggregations of data at the researcher level (Leahey et al., 2017), and we test the individual incentives in pursuing IDR. We measure the scholars' reputation by looking at the number of citations accrued by articles and their contribution to research with societal impact through the articles' degree of generality. Interdisciplinarity has so far been intended uniquely as a way of combining different sources of knowledge, but, it is our conviction, that it is circulation of such knowledge that realizes the public interest associated with IDR. As societies become more interconnected and grow in complexity, science needs to combine knowledge from different domains but also shares new findings with them. Following Carley and Porter (2012) and Fontana et al. (2020), we measure generality by calculating the dispersion of citations across disciplines through the Hirschman-Herfindahl Index.[3]

We measure interdisciplinarity by highlighting its main dimensions (Porter and Rafols, 2009; Yegros-Yegros et al., 2015): the number of fields embedded in a paper (Variety); the evenness of their distribution (Balance), and the similarity between

[1]The University of Florida is a large research university in the United States that comprises more than 5,000 researchers and 50,000 students. UF consistently ranks among the top ten public universities in the United States and is the flagship university in the state of Florida.

[2]In principle, also other datasets such as MAG may allow creating a longitudinal dataset about scholars using an identification code. However, such identification codes are obtained through inferential methods, and they are not directly registered by scholars or their institutions. On the contrary, our information is more reliable since the association of articles to the same scholars is done by the UF, and there is no inference involved.

[3]The index is widely applied in economics of innovation literature to measure the range of inventions that derive from a patent (Bresnahan and Trajtenberg, 1995; Squicciarini et al., 2013).

them (Disparity).[4] The use of multiple and distinct indicators allows capturing all the facets of a complex concept like interdisciplinarity.

Our identification strategy relies primarily on the use of individual, disciplinary-based citation patterns and year fixed effects, which allow registering the effect of a change in interdisciplinarity on the scientific impact of a researcher while sorting out potential confounding factors and the influence of a change in other dimensions. The additional information contained in our database, moreover, give us the chance to shed light on different sources of heterogeneity and assess whether the impact of IDR differs across gender, collaboration types, research proficiency, and disciplinary affiliation.

Our findings confirm the existence of a trade-off between private and public interests in one of the three observed interdisciplinarity dimensions. An increase in the evenness of the distribution of disciplines in article references (Balance) results in a decrease of the number of accrued citations, but increases its generality. In addition, we find that the increase of the number of disciplines recombined in a paper (Variety) has a positive effect on the number of citations and generality. This seems to signal a private incentive to and public benefit from pursuing IDR, however, when the distance of the involved disciplines (Disparity) increases both citations and generality decline. Therefore, a trade-off emerges, independently of the involved interests, among the dimensions of IDR. Importantly, results are confirmed even when considering scholars with different characteristics or affiliations. In other words, all scholars face the similar incentives and constraints in engaging in interdisciplinary projects.

This evidence suggests that much effort should be put into coordinating private and public interests by tuning hiring and rewarding mechanisms with funding policy whenever interdisciplinarity is concerned. Secondly, the private and public benefits of IDR do not grow infinitely: in spite of its undeniable importance, interdisciplinarity is not the panacea for all scientific and societal issues.

This paper aims to make three contributions. Firstly, we provide evidence on the existence of a trade-off between private (researchers) and public (society) benefits in pursuing IDR. Secondly, we propose a novel approach to analyze researchers' scientific outcomes from a micro perspective without the aggregation of bibliometric data. Finally, we introduce the generality of knowledge as an additional measure of interdisciplinarity and, at the same time, as a relevant indicator of the achievement

[4] The literature also uses the Rao-Stirling diversity (Stirling, 2007), an index that synthesizes the three dimensions. In addition to the loss of details, it has been shown (Fontana et al., 2020, Figure 10) that the Rao-Stirling diversity is highly correlated with Disparity. We, therefore, decided not to include it in our analysis.

of the public goal to obtain interdisciplinary solutions to societal problems.

The paper proceeds as follows. Section 1.2 presents the theoretical background and motivating evidence, while Section 1.3 summarizes our research hypotheses. In Sections 1.4 and 1.5, we describe our empirical strategy and data, respectively. Section 1.6 discusses the results, and Section 1.7 concludes.

## 1.2 Interdisciplinary and researchers' incentives

A vast and growing literature has stressed the existence of *multiple logics* within the academia (Llopis et al., 2022): researchers might engage in activities that pursue rather different goals. They can engage in quasi-market action such as academic patenting, academic entrepreneurship (Sterzi et al., 2019), they can act to increase their reputation within academia, and finally, they can endeavor towards research with a higher societal impact (Mazzucato, 2018).

It has been convincingly argued by Llopis et al. (2022) that the multiplicity of objectives can make it difficult for researchers to respond to conflicting incentives and that policies that sustain different logics might aggravate the issue. Several studies have explored the trade-off between market and scientific activities (see, for instance, Tartari and Breschi, 2012), while the individual and institutional tension between reputation building and societal activities remains unexplored.

In this paper, we adhere to the definition of reputation proposed by Llopis et al. (2022, p. 2): the scientist's academic status within her peer community. Reputation gives scientists recognition and leverage in competitions and funding. We assume that such status is mainly built through the publication of articles (Subramanian et al., 2013) and their subsequent citations (Hamermesh and Pfann, 2012; Jamali et al., 2016; Jones, 2021). Instead, we define societal research as the activity that tackles issues that are "complex, systemic, interconnected, and urgent, requiring insights from many perspectives" (Mazzucato, 2018, p. 803). We assume that given its nature, societal research primarily requires insights from different perspectives and the subsequent circulation of the derived knowledge beyond disciplinary fields (The National Academies, 2005). We then ask if scientists can simultaneously achieve reputation – recognition in their own field – and contribute to societal research. We use the interdisciplinarity of scholars to provide them with a degree of involvement in reputation-seeking behavior and societal research.

The existing literature on IDR primarily focuses on scholars' scientific outcomes, rather than researchers themselves (Leahey and Barringer, 2020; Hackett et al., 2021), and therefore is only partially relevant to this study. Several studies highlight the mixed effect of the various aspects of interdisciplinarity on citations, measured as the number of citations received by single articles (see, among others, Fontana et al., 2020; Yegros-Yegros et al., 2015).[5] Results vary across the dimensions of IDR and disciplines taken into account. However, those studies commonly identified an inverted U-shaped relationship between the interdisciplinarity and impact of an article. Moving from articles to research projects and grants, Bromham et al. (2016) suggested the existence of a bias against interdisciplinarity in funding evaluations.

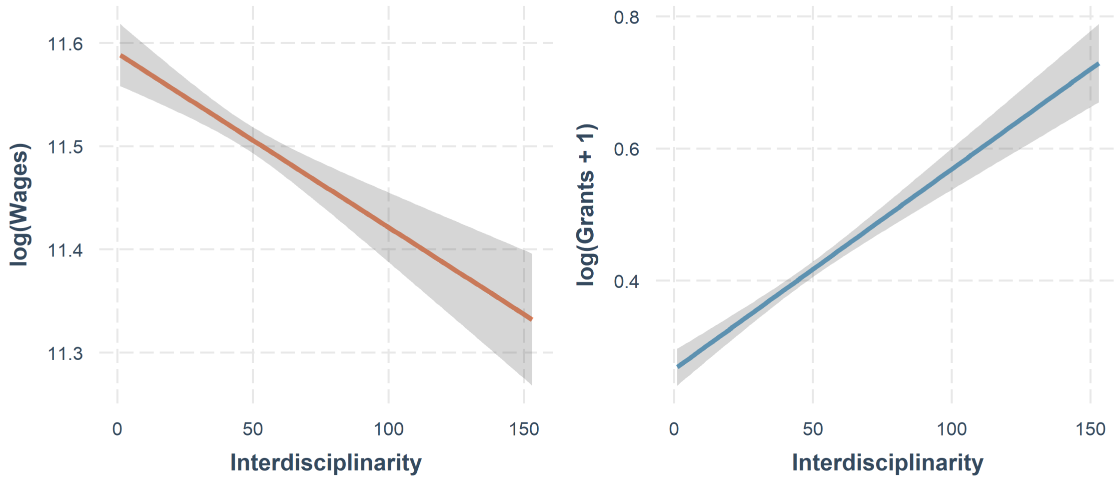
While the effect of interdisciplinary on knowledge production and scientific impact has been extensively studied in the literature, the impact of pursuing IDR on scholars' productivity, career, and funding performance is still underexplored. The existing evidence, however, seems to highlight that IDR comes with a cost. Leahey et al. (2017) provided one of the first studies on potential scholars' costs and benefits associated with interdisciplinarity research. They collected 32,000 articles published by 854 researchers from a wide range of fields and universities. The authors computed researcher-level bibliometric indicators by considering scholars' publications in the entire period of analysis. Overall, they found that an increase in the average interdisciplinarity of scholars' work improves their visibility in the scientific community, measured as the cumulative number of citations, and decreases their productivity as indicated by the number of articles published. Sun et al. (2021) analyze 44,419 research grant awarded by the research councils in the UK and find that interdisciplinary research is less impactful than specialized research in the short run but, eventually, is more rewarding in terms of volume and value of funding.

We are supported in our research questions by a preliminary evidence on the effect of interdisciplinarity on yearly wages and research funding (number of grants) in a sub-sample of scholars at the University of Florida in the period 2008–2013 (more details about data are in Section 1.5). Once controlled for scholars' academic age, we use their wages to represent a signal of academic reputation and the number of awarded grants to indicate their potential contribution to societal research. Then, we create an interdisciplinary profile of researchers, i.e. the extent to which they are prone to conduct interdisciplinary research, by using the maximum number of unique disciplines in the references of an article written by an investigator in one year.

Figure 1.1 shows that there exists a negative and statistically significant correlation between interdisciplinary and researchers' wage, while we observe a positive and statistically significant correlation between interdisciplinary and the number of

[5] For a survey of the literature on interdisciplinarity see Wagner et al. (2011), for a review on the relationship between interdisciplinarity and impact see Zeng et al. (2017, section 6.1.1).

Figure 1.1: Correlation between interdisciplinarity and academic achievement




The regression line represents the correlation between the interdisciplinarity profile of a sample of researchers at UF and their academic achivements: the yearly wage (left-side panel) and the number of grants obtained in a year (right-side panel). Full results in 1.14.

grants received by researchers. In other words, scholars that conduct research in delimited fields of study receive higher wages, while more interdisciplinary researchers are awarded with more grants. This evidence thus corroborates our hypothesis that researchers receive contrasting incentives when engaging in interdisciplinary work. By increasing the interdisciplinary content of their research, scholars also increase their societal relevance and receive more grants. At the same time, this reduces scholars' reputation within their academic circle resulting in lower wages.

To confirm and better understand the mechanisms that lead to these contrasting outcomes, in the following sections, we will investigate the reasons behind the observed difference by looking at the main drivers of reputation building and societal impact. Namely, keeping all other variables constant, we will focus on the number of citations accrued by a scholar as one important evaluation criteria in career progressions and therefore in the wage level. We then look at the diffusion of the knowledge across disciplines as the fulfillment of the interdisciplinarity required by funding agencies. It is worth anticipating that, while retaining the scholar's perspective, we will perform our analysis starting from papers. This allows us to characterize scholars' research at a more fine-grained level than what is allowed by variables that concern scholars. Moreover, by considering articles and not aggregating data at the researcher level, we are able to distinguish among the different dimensions of IDR.

## 1.3 IDR and researchers' trade-offs: research hypotheses

To capture the different facets of IDR that might influence research' scientific outcomes, we measure interdisciplinarity as the Diversity of the combined knowledge, i.e. "the apportioning of elements or options in any system" (Stirling, 2007). In fact, several mechanisms exist through which IDR might affect scientific impact, and the existence and extent of the supposed trade-off between private and public benefits might also vary considerably across the dimensions of IDR. We rely on the literature that decomposes Diversity in three independent components (Fontana et al., 2020; Hackett et al., 2021; Porter and Rafols, 2009; Stirling, 2007; Yegros-Yegros et al., 2015), defined at the article level: Variety, Balance, and Disparity.[6] The three dimensions of Diversity have specific meanings and autonomy, and refer respectively to the number of different disciplines involved in the making of the paper, their relative frequency, and their distance.

Variety is the basic form of interdisciplinarity: it returns the number of different disciplines that are referenced in the paper. It provides *prima facie* evidence on the intensity of interdisciplinarity of an article, but gives no information on the relative importance of the involved disciplines.

Balance overcomes this drawback by building on Variety in order to quantify the distribution of disciplines in the article references. Namely, it measures the evenness of the distribution of disciplines in references. Low values of Balance indicate that the paper references articles from a prevailing discipline, while high values of Balance correspond to an even distribution of disciplines in references.

Disparity measures a further dimension of Diversity: the proximity of the referenced disciplines in the knowledge space. The underlying idea is that disciplines that frequently co-occur in references are closer than those that co-occur rarely with respect to all other occurrences. High values of Disparity signal that a paper references fields that are very distant – have a low proximity – in the knowledge space. This indicator is rather different from Variety and Balance in that it does not heavily depend on the system of data classification as they do: proximity is calculated over the entire sample of articles and, therefore, provides the effective relative distance between pairs of disciplines. We will provide further details on the operalization of these indicators in Section 1.4.

The channels through which the IDR dimensions can affect the reputation and

[6] Diversity also includes a compound indicator, the Rao-Stirling diversity, that is more suitably computed when the distinct role of the IDR components is not relevant to the object of analysis.

the societal contribution of a scholar are diverse. Firstly, there might exist a trade-off between the different dimensions of IDR. Increasing Variety implies that the pool of possible citing scholars increases. As a result, this component of IDR might positively impact both the number of citations and the diffusion of knowledge across fields. However, this might not hold when the referenced disciplines are very distant to one another or when the focal paper is hardly identifiable with a field of study and, likely, less useful for a wide range of disciplines. This results in a trade-off for the researcher that pursues IDR, since increasing Variety will eventually end up in increasing Disparity.

Moreover, while the increase in some components of IDR is likely to positively affect the circulation of knowledge (public benefit), it might penalize the scholar prestige in a highly specialized academic environment (private benefit). This aspect might be particularly relevant for Balance: an even distribution of references to different disciplines may encourage the diffusion of the paper across a wide range of fields, but, at the same time, the paper will not have a target scientific community and will hardly be highly cited.

Combining these insights, we developed the first two hypotheses that we will test in our empirical analysis:

**Hypothesis 1a (HP1a)**: *If IDR has an effect on the scholars' reputation, this impact differs across the various dimensions of IDR: while high Variety increases the potential to be cited by a larger set of scholars, a growth in Balance and Disparity might reduce the number of citations received by an article, since it will hardly fit within a defined field of study. Therefore, a trade-off in scholars' private benefits exists.*

**Hypothesis 1b (HP1b)**: *If IDR has an effect on the circulation of knowledge, this impact differs across the various dimensions of IDR: while high Variety and Balance increases the potential diffusion of knowledge across fields, a growth in Disparity might reduce the circulation of an article across disciplines, since it will be more difficult to integrate in the existing literature. Therefore, a trade-off in public benefits exists.*

The different impacts of IDR components on scientific outcomes also result in a trade-off between private and public benefits. In this respect, we will test the following hypothesis:

**Hypothesis 2 (HP2)**: *If IDR has an impact on the scholars' reputation and circulation of knowledge, the effect differs across these two indicators of scientific outcome: the increase in the Balance in IDR hampers receiving a high number of citations, but favors knowledge diffusion across disciplines. Therefore, a trade-off* *between public and private benefits in pursuing IDR exists.*

## 1.4 Empirical strategy

The aim of our empirical analysis is to compare articles with different interdisciplinary content and assess whether they have a different scientific impact.

In order to make sure that articles are fairly compared, we elaborate an empirical design which allows us to compare only articles with similar characteristics, but with a different interdisciplinary content, published by the same author within the same field of study during the same year. Of course, interdisciplinarity is only one of the many factors determining the scientific impact of an article. If these factors are not considered, we would have a problem of omitted variables biasing our analysis. For this reason, we make sure that comparison is conducted sorting out specific features of the article and time-varying characteristics of the author which may concur to explain the scientific impact of an article.

In practice, our analysis is conducted using the following model:[7]

$$Y_{ijft} = IDR_{ijft}\beta + X_{it}\gamma + K_{jf}\delta + \alpha_i + \phi_f + \theta_t + \epsilon_{ijft}. \tag{1.1}$$

Here, the dependent variable ($Y_{ij}$) is a measure of the scientific impact of paper $j$ written by investigator $i$ at time $t$ in the field of study $f$, measured alternatively as the number of citations of paper $j$ or its generality index (see Section 1.4.1 for their definitions), and the regressor of interest is $IDR_{ijft}$, which measures the various interdisciplinarity dimensions of paper $j$ as defined in Section 1.4.1 (i.e. Variety, Balance, and Disparity).

The variables $\phi_f$, $\theta_t$, and $\alpha_i$ denote fields of study, year, and investigator fixed effects, respectively. These allow to compare only articles with similar characteristics, considering different sources of unobserved heterogeneity which may interfere with the effect that interdisciplinarity has on the scientific impact of an article: i.e., time-invariant characteristics of the article's field of study, publication year, and author. The variables $K_{jf}$ and $X_{it}$ are a proxy of the characteristics of the article and the author, respectively. They sort out potential problems of omitted variables in the model specification by controlling for specific features of the article (i.e., the number of authors, the presence of collaborators affiliated to an institution outside

[7]Estimates are obtained using an ordinary least squares regression. For the model specification where the dependent variable is the number of citations, we test the robustness of our results to the choice of the estimator. Specifically, we estimate our model using both Poisson and Negative Binomial regressions. Results are qualitatively unchanged. They are presented in Table 1.15 in Appendix 1.8.4.

the United States, the adoption of a monodisciplinary approach)[8] and time-varying characteristics of the author (i.e., the H-index of investigator $i$ at time $t$, that is an author-level metric that measures cumulative productivity and citation impact of the researcher) which may concur to explain the scientific impact of an article. In order to avoid over-weighting extreme values in our estimates, and correctly deal with the highly skewed nature of our continuous variables, these are all log-transformed. The descriptive statistics for these variables in our data are presented in Table 1.1.

In the model, the parameter of interest is $\beta$, i.e., the estimated coefficient associated to $IDR_{ijft}$. This has to be interpreted as the average effect of an increase in the interdisciplinary content of an article on its scientific impact, *all else being equal*: i.e., when comparing articles with similar characteristics, but with a different interdisciplinary content, published by the same author within the same field of study during the same year. The robustness of our estimates relies on the fact that we are able to sort out from the model any identification threat arising from the presence of omitted variables (i.e., specific features of the article, $K_{jf}$, and time-varying characteristics of the author, $X_{it}$), and from performing unfair comparisons between articles due to potential unobserved heterogeneity in our data (i.e. time-invariant characteristics of the author, $\alpha_i$, the field of the article, $\phi_f$, and the year in which the article was published, $\theta_t$). Importantly, the estimated value of $\beta$ can be considered as representative of a large population, since our data covers a large number of authors working in many different fields across several years.

It is important to stress that we can rely on this sound empirical design because of our rich and innovative source of data which keeps track of the career of researchers working in several disciplines over different years, and allow us to use individual, field of study, and time fixed effects. To the best of our knowledge, we are the first to use this model specification in this strand of research.

### 1.4.1 Interdisciplinarity and scientific outcome indicators

As anticipated in Section 1.3, following Stirling (2007), we define three different dimensions of interdisciplinarity: Variety, Balance, and Disparity. We compute these indicators by using the disciplines of the papers listed in the references of the focal articles.

Variety measures the number of different disciplines referenced by the paper.

[8] Please observe that in these cases Balance and Disparity are not defined.

Thus, we define Variety ($V_j$) as:

$$V_j \equiv \sum_{s \in F} 1, \tag{1.2}$$

where $F$ is the set of disciplines $s$ in references of a paper $j$.

Balance, instead, refers to the evenness of the distribution of disciplines. We operationalize the Balance ($B_j$) as a normalized Shannon Entropy, defined as:

$$B_j \equiv \frac{1}{\log V_j} \sum_{s \in F} f_s \log f_s, \tag{1.3}$$

where $V_j$ is the Variety measured as above and $f_s$ is the frequency of discipline $s$ in references of paper $j$. After normalization, this index assumes values between 0 and 1.

Finally, Disparity ($D_j$), which concerns the distance among referenced disciplines, is defined as the normalized sum of proximity among fields:

$$D_j \equiv \frac{1}{V_j(V_j - 1)} \sum_{\substack{r,s \in F \\ r \neq s}} (1 - p_{rs}), \tag{1.4}$$

where $p_{rs}$ is the proximity between disciplines r and s. The computation of proximity is usually based on the co-occurrence of disciplines in articles, normalized by the size of fields. A common indicator is cosine similarity, which measures the cosine between fields' vectors of co-occurrences in references. Disparity is defined for values between 0 and 1 and is independent of Variety and Balance. It is worth noting that Balance and Disparity are not defined for articles that cite only one discipline (i.e. when Variety is equal to one).

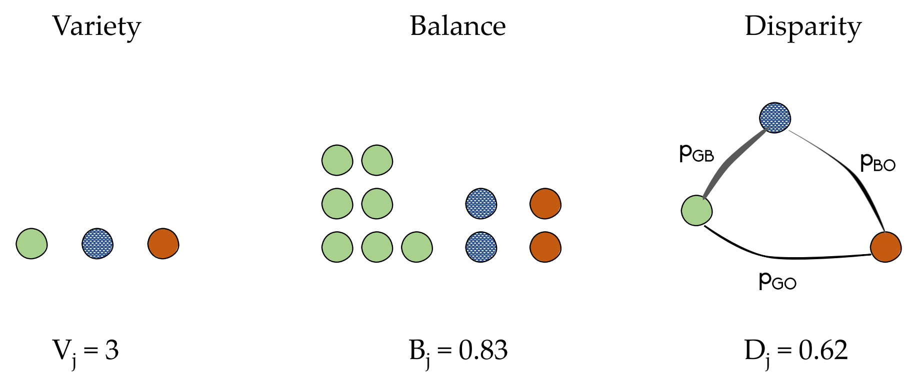
Figure 1.2 exemplifies the three measures of interdisciplinarity in the case of a paper that cites three unevenly-distributed disciplines, with different proximity to each other.

For what concerns the scientific impact, we operationalize researchers' reputation in academia as the total number of citations received in a five-year period after the publication date (Hamermesh and Pfann, 2012). It is described as:

$$C_j \equiv \sum_{t=y_{pub}}^{y_{pub}+5} c_{jt}, \tag{1.5}$$

where $y_{pub}$ is the article's publication year and $c_{jt}$ represents the citations received by a paper $j$ in year $t$. We count citations over a five-year time window to have an

Figure 1.2: Example to illustrate the IDR measures.




The example article cites three different disciplines (Green, Blue, Orange), with a prevalence of Green (7) over Blue (2) and Orange (2). In Disparity, the strength of links between fields of study is proportional to their mutual proximity. In this example, Green and Blue are similar to each other (they are often cited together, i.e. they frequently co-occur in references), while Orange is more distant.

indicator that is consistent between papers published in different years.

To measure knowledge diffusion across disciplines, instead, we rely on an index of generality of knowledge. A bit of knowledge that influences many, possibly distant, disciplines can be thought of as more impactful than one that is received only by few disciplines (Carley and Porter, 2012). This index captures the degree of applicability and influence of knowledge of paper on different fields of study. It is computed using the Hirschman-Herfindahl concentration index of citations across disciplines (Hall et al., 2001; Trajtenberg et al., 1997) and is defined as:

$$G_j \equiv 1 - \sum_{f=1}^{|F|} \frac{N_{jf}^2}{N_j^2}, \tag{1.6}$$

where $N_{jf}$ is the number of forward citations received by a paper $j$ from papers in the field of study $f$, while $N_j$ is instead the total number of forward citations received by the paper. By definition, Generality is bounded between 0 and 1. Articles having their citations spread among many disciplines will have a high value of this indicator.[9]

[9] One shortcoming of this measure is that it is not defined in articles that did not receive any

## 1.5 Data

We construct a novel and unique dataset that includes detailed information about researchers and their publications: we study all the researchers affiliated to the University of Florida in the period 2008-2013. UF is the flagship university in the state of Florida, it is a large research university comprising more than 50,000 students and 5,000 full-time faculty. Over the past ten years, research awards to the university have increased by 45%: from $619 million in 2011 to 900.7 million in 2020.[10] UF is a member of the Association of American Universities, an organization of sixty-two academically prominent public and private research universities in the United States and Canada, and it consistently ranks among the top ten public universities in the United States. UF therefore represents an excellent example of a prominent and large research-oriented institution, and for this reason it has been already used as a case study to investigate how scientific collaborations are formed (Leone Sciabolazza et al., 2017), the mechanisms of scientific team assembly (Smith et al., 2021), and the design of new research policies (Leone Sciabolazza et al., 2020).

From the UF's registry office, we obtained information on researchers' gender, department affiliation, and publication record.[11] The individual-level information is anonymous, thus researchers' names are substituted by a unique identifier. The investigators' publication records provided by the UF's registry office include articles' title, journal in which the article was published, and the publication year. We exploit the publication title to retrieve the Digital Object Identifier (DOI) assigned to each article, i.e. the unique identifier of the publication in all bibliometric databases.[12] Through DOIs, we then collect articles' citations and references from the Lens database, while papers' fields of studies and authors' institutional affiliations were collected from the Microsoft Academic Graph (MAG) database.[13] We use information about citations received by papers to compute both scientific impact indicators and researchers' H-index, which will be our proxy for the quality of

citations in the five-year windows. This may lead to selection bias concerns that are discussed in the following sections.

[10]From: *University of Florida hits record $900 million in research awards*, University of Florida News (2020). Available at: https://news.ufl.edu/2020/08/record-research-awards/.

[11]We focus on articles published in peer-reviewed journals, excluding books and other types of academic production from our analysis.

[12]This process exploits Crossref and Scopus APIs. The search procedure is described in Appendix 1.8.1.

[13]The Lens database and Microsoft Academic Graph database used to complement information on articles by UF's researchers are becoming widespread for bibliometric analysis in recent years. Given that the fields of study information are crucial for our IDR measures, we decide to rely on these sources to maintain consistency and uniformity between our databases. Both sources can be freely accessed for research purposes and available at the following links: Microsoft Academic Graph and Lens.

scholars.[14]

To determine disciplines associated to articles and compute interdisciplinarity indicators, we rely on the classification scheme implemented by MAG to retrieve the field of studies associated to each paper. This scheme is a hierarchical classification that identifies 19 disciplines (first level) and 292 sub-disciplines (second level). The taxonomy uses state-of-the-art artificial intelligence methodologies to extract semantic content from documents, exploring natural language processing techniques and networks semantic reasoning to delineate disciplines (Sinha et al., 2015; Wang et al., 2019). There are several advantages in using this classification: it is based on concepts and language used at the paper-level, thus it avoids any bias that may arise from arbitrariness in the details of classifications that rely on human experts (Wang and Schneider, 2020);[15] it uses a heterogeneous network semantics analysis that exploits the context in which the publication's text is embedded, linking it to authors, affiliations, and locations (Wang et al., 2019); and it also mitigates the assignment errors that results from the loss of granularity when we adopt journal-based categorization. Moreover, journal-based taxonomies have difficulties in dealing with generalists journals like *Nature*, *Science*, and *PLoS ONE*.

In our final database, we observe 6,105 researchers at UF, of which 34% are women, with at least one article in a peer-reviewed journal in the period 2008-2013. On average, the period of activity of each scholar in our sample (i.e. the number of years in which she publishes at least one journal article) is three years. At UF, researchers belongs to different colleges, which, in turns, are aggregated in four academic units: Liberal Arts and Sciences, Engineering, Health Sciences, and Food and Agricultural Sciences. Scholars in Health Sciences, especially in the college of Medicine, prevail in our sample (more details in Table 1.9). In addition to these pieces of information, the UF's registry office also reports yearly wages and number of awarded grants for a limited number of researchers. These data have been exploited in the motivating evidence (see Section 1.2) and are described in Table 1.13.

The full publication record of UF's researchers consists of 23,926 articles published in peer-reviewed journals. As reported in Figure 1.5, the number of publications by year is quite stable over time, with about 4,000 articles per year. Overall, these papers made 646,280 references and received 366,024 citations in five years

[14] The H-index is an author-level metric that measures cumulative productivity and citation impact of each researcher. It takes into account the scholar's best cited papers and their number of citations. A researcher with $n$ papers with at least $n$ citations will have a H-index of $n$.

[15] For example, the total number of categories of the two most frequently used system of classification, Web of Science (WoS) journal subject categories (SC) and the All Science Journal Classification (ASJC) from Scopus, varies drastically: there are 252 SCs and 330 ASJCs.

Table 1.1: Summary statistics.

| **Variables** | *Mean* | *SD* | *Min* | *50%* | *Max* | *Obs* |
|---|---|---|---|---|---|---|
| **Panel A: Researcher-level Data** | | | | | | |
| Nb. papers/year | 2.22 | 2.08 | 1 | 1.6 | 40.33 | 6,105 |
| Nb. citations/year | 17.65 | 42.37 | 0 | 5.0 | 955.50 | 6,105 |
| H-index | 2.37 | 2.88 | 0 | 1.5 | 35.83 | 6,105 |
| Gender (Woman=1) | 0.34 | 0.48 | 0 | 0 | 1.00 | 6,105 |
| **Panel B: Paper-level Data** | | | | | | |
| Nb. Citations | 20.30 | 46.34 | 0 | 10 | 2,530 | 23,926 |
| Generality | 0.72 | 0.18 | 0 | 0.77 | 0.98 | 22,658 |
| Variety | 37.06 | 19.54 | 1 | 36 | 153 | 23,926 |
| Balance | 0.84 | 0.09 | 0 | 0.85 | 1 | 23,926 |
| Disparity | 0.68 | 0.07 | 0 | 0.70 | 0.94 | 23,926 |
| Nb. References | 40.21 | 33.01 | 1 | 34.00 | 926 | 23,926 |
| Nb. of Authors | 5.64 | 9.90 | 1 | 4 | 1,269 | 23,926 |
| International Collab. | 0.23 | 0.42 | 0 | 0 | 1 | 23,926 |

*Notes*: Panel A shows selected measures of productivity of 6105 researchers affiliated to the University of Florida from year 2008 to 2013. Gender is a dummy variable that assumes the value 1 when the researcher is a woman. Panel B shows descriptive statistics of the 23,926 articles published by these researchers in the time window 2008-2013. Nb. Citations is the total number of citations received in a 5 years time after the publication. The Generality captures the degree of applicability of the knowledge codified in a paper on different fields of study. It is worth noting that generality is not defined for papers with zero citations. International collaboration is a dummy variable that assumes the value 1 when at least one co-author in the paper is affiliated to an institution outside the United States.

from the publication date. Considering only the years of activity, each UF's researcher published an average of 2.22 papers per year. 23% of these papers involves international collaborations, and 46% of them has more than one UF's researcher as an author. More details about researchers' and articles' characteristics are in Table 1.1.

Each article in the sample belongs to one or more disciplines, as measured by the MAG field of study classification. We exploit the most fine-grained level of this hierarchical classification (second level) to define articles' degree of interdisciplinarity and generality (see Section 1.4.1 for the definition of indicators). This level consists of 292 categories, that can be aggregated in 19 more general fields of study (first-level classification). While the second level of classification is used to compute all article-level indicators, we consider the first level of classification to define discipline fixed effects and, thus, control for different citation patterns across disciplines. In this section, we refer to these 19 fields of study at the first level of classification also for descriptive purpose, in order to describe articles' characteristics. The distribution of papers over these 19 categories is reported in Table 1.2. As expected, the average number of references and the average number of citations is heterogeneous across fields of study. More details about the number of references and citations by discipline are available in Table 1.10.

Table 1.2: Distribution of focal papers by field of study (first level of classification).

| **Field of Study** | **Total** | **Average Nb. References** | **Average Nb. Citations** |
|---|---|---|---|
| Biology | 7781 | 46.07 | 22.26 |
| Medicine | 6305 | 35.64 | 22.14 |
| Chemistry | 2628 | 41.60 | 19.86 |
| Psychology | 1703 | 46.51 | 17.31 |
| Physics | 1686 | 39.27 | 22.71 |
| Mathematics | 996 | 26.48 | 9.87 |
| Materials science | 785 | 34.34 | 21.68 |
| Computer science | 508 | 31.73 | 12.65 |
| Geology | 506 | 48.77 | 17.96 |
| Economics | 503 | 37.41 | 13.82 |
| Engineering | 404 | 27.78 | 13.71 |
| Sociology | 199 | 34.58 | 8.27 |
| Environmental science | 85 | 39.41 | 52.58 |
| Geography | 59 | 44.47 | 29.29 |
| History | 39 | 29.44 | 3.36 |
| Political science | 29 | 26.90 | 11.10 |
| Business | 26 | 57.19 | 24.00 |
| Philosophy | 20 | 31.75 | 2.15 |
| Art | 7 | 11.29 | 1.57 |

*Notes*: This table shows the distribution of focal articles per fields of study at the first level (19 categories). The average number of references relates to papers cited by our articles of interest and the average number of citations takes into account total number of citations within 5 years from the publication.

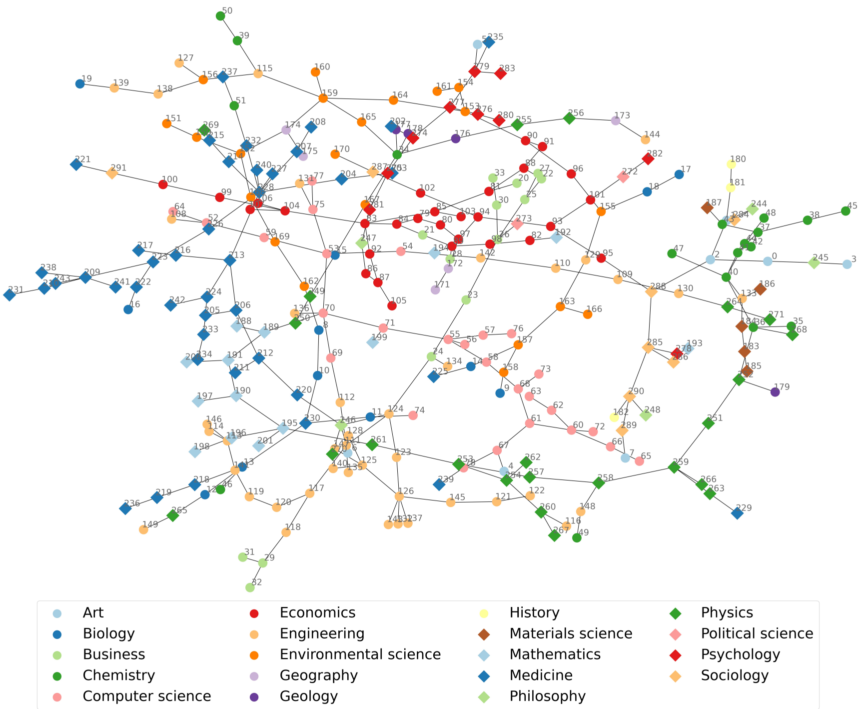
The information about MAG fields of study is also used to compute the knowledge space in which the scholars perform their research. The knowledge space, which summarizes the proximity between disciplines, is the core of the Diversity indicator, one of the dimension of IDR considered in this paper. As we are interested in fine-grained definitions of IDR indicators, we consider the knowledge space at the second level of discipline classification (292 fields of study). To avoid biases due to the small number of papers in our sample and obtain a more reliable measure of similarity between disciplines (as a proxy of the easiness in combining different topics and techniques in a single research), we use an index of proximity among fields of study computed over the universe of articles in MAG. This proximity measure is based on the Network Similarity Package, a series of processing functionalities for MAG that allow us to compare two fields of study and obtain a similarity score that represents how close these fields are, based on the frequency they appear together in a same paper.[16] Based on this measure of similarity, we represent the network of fields of studies, i.e. the knowledge space, in Figure 1.3. The graph connect disciplines whose co-occurrence is frequent in the universe of MAG articles. Nodes represent fields of study at the second level of MAG classification, but, to ease the interpretation of the knowledge space, their shapes and colors correspond to disciplines at the upper level of classification (conversion table is available in Appendix 1.8.5). In the graph, sub-disciplines belonging to environmental science, medicine, and biology are on the left. At their right, we can observe the interconnection between economics and business. The bottom part of the network, instead, shows the connection between fields in mathematics (starting from the left), engineering, computer science, chemistry, physics, and material science. At the top of the figure, the interpenetration between art (included literature), psychology, sociology, history, and geography is evident.

Beyond the information about the relative distance between disciplines, the field of study classification and the knowledge space allow us to define the three different dimensions of IDR in our sample, as explained in Section 1.4.1. Figure 1.4 shows average values of Variety, Balance, and Disparity by field of study (at the first level of classification). While the average values of these indicators do not differ considerably across disciplines, some fields of study have unique characteristics in terms of interdisciplinarity. The most evident one is Art, as it has the lowest average Variety and Disparity and the highest Balance in the sample. Those values characterize Art as a poorly diversified discipline, in which, however, different fields are evenly combined in article references. The opposite occurs in Business. In this field of study, the articles show, on average, a high Variety and Disparity – meaning that they are highly diversified – but a relatively low Balance – signalling the presence of a core field in article references. The importance of a core field of study (low Balance) is especially relevant in Philosophy, Biology, and Physics. History, instead, results as a highly specialized field since it has a relatively low value in all three indicators.

In the following section, we explore in more details the different dimensions of

[16] For details on the Network Similarity package, see Microsoft Research (2020).

Figure 1.3: Knowledge space among fields of studies.




The network shows the proximity between fields of study at the second level of the MAG classification (292 fields of studies). To ease the graph' interpretation, authors grouped fields of studies by discipline (the first level of MAG classification), which are represented by different colors and shapes, as reported in the plot legend. The conversion between the two levels as well as the field of studies corresponding to node IDs are reported in Table 1.17.

interdisciplinarity by using the publication records of UF's researchers described in this section. The design of our empirical strategy requires both article-level and researcher-level information. By matching article-level and individual-level information for each UF's author of our papers, we obtain 46,156 observations at the paper-researcher level as co-authorship between UF's researchers is frequent in our sample. Descriptive statistics at the paper-researcher level are available in Table 1.11.

## 1.6 Results

### 1.6.1 Main results

In this section, we present the results from the estimation of equation (1.1) to assess the average effect of an increase in the interdisciplinary content of an article on its scientific impact, *all else being equal*: i.e., when comparing the scientific

Figure 1.4: Average Interdisciplinarity by field of study


The colored columns represents the average variety, balance, and disparity by field of study at the first level of classification. The color of the fields represents the researchers' corresponding network measures - bluer dots represent lower values while redder dots represent higher values.

impact of articles with similar characteristics, but with a different interdisciplinary content, published by the same author within the same field of study during the same year. Findings from this exercise will be used to investigate the evidence in favor of hypotheses HP1a, HP1b, and HP2.

HP1a and HP1b posit that the impact of IDR on a given measure of scientific outcome differs according to the interdisciplinarity dimension considered. In order to test these hypotheses, we will assess whether the different dimensions of interdisciplinarity have a similar effect in determining the scientific impact of a paper (i.e. either citations or generality), or some of them are considered desirable and are rewarded by the academia while others are less desirable and thus penalized. Specifically, H1a states that, while Variety has a positive influence on the number of citations, the opposite occurs with Balance and Disparity. If the latter is verified, we would find evidence of the existence of a trade-off in researchers' public benefits. H1b, instead, conjectures a positive effect of Variety and Balance on the diffusion of knowledge and a negative impact of Disparity on the same scientific outcome. In this case, we expect to find evidence on the presence of a trade-off in the public interest and societal benefit.

HP2 states that the impact of a given dimension of IDR differs according to the measure of scientific impact considered: while we expect to observe a negative effect of Balance on the number of citations, a positive impact of the same IDR dimension is supposed to have a positive influence on the generality of knowledge. In order to test this hypothesis, we will investigate whether the same dimension of interdisciplinarity (Balance) has the same impact when considering different measures of scientific impact, i.e. citations and generality, or this is rewarded in some cases and penalized in other cases. If evidence supports the latter scenario, then results would confirm our hypothesis and the existence of a trade-off between private and public benefits in pursing IDR.

We begin our investigation by testing HP1a. To this purpose, we assess the effect of an increase in the interdisciplinary content of an article on its scientific impact when this is measured in terms of number of citations (reputation).[17] Results are reported in Table 1.3. In column (1), we jointly estimate the effects of the three dimensions of interdisciplinarity, so to assess the effect of an increase in the interdisciplinary content of a paper in one dimension (e.g., Variety), while accounting for changes in other interdisciplinary dimensions (e.g. Balance and Disparity). In order to consider different potential sources of unobserved heterogeneity which may interfere with the effect that interdisciplinarity has on the scientific impact of an article, we include monodisciplinarity, individual, field of study, and year fixed effects into our model specification. We find that only an increase in the Variety of a paper has a positive and statistically significant effect on its number of citations, whereas the other two measures (Balance and Disparity) have a negative and statistically significant effect. Our results are thus in favor of HP1a, and we observe a trade-off in scholars' private interest.

In column (2), we augment our model specification by including a control for the number of authors in the paper. The estimated coefficient of this variable indicates that increasing the number of authors has a positive and statistically significant impact on the number of citations of a paper. This is consistent with the idea

[17]It is worth noting that we compute the interdisciplinary indicators by considering the second level of discipline classification (292 fields of study), while we use the first level of classification (19 fields of study) to define discipline fixed effects that controls for the presence of different citation patterns across disciplines.

that the narrower expertise of researchers requires having larger teams to producing widely-cited research. Most importantly, all our previous findings in favor of HP1a are confirmed: i.e., the sign and the statistical significance of the three dimensions of interdisciplinarity are unchanged with respect to column (1).

In column (3), we add to our model specification a dummy variable registering whether one of the co-authors of the paper is affiliated to an institution outside the US. We find that having an international collaborator in the team has a positive and statistically significant effect on the number of citations, hinting that working in an international team may expand the visibility of one's work. Again, all our results supporting HP1a are left qualitatively unchanged.

In column (4), we add a control for the H-index of the investigator, thus estimating equation (1.1) with its entire set of controls. We find that having a higher H-index has a positive and statistically significant effect on the number of citations. Most relevant to us, the evidence in favor of HP1a is still confirmed.[18] Even after including the entire set of controls in our model specification, interdisciplinarity has a large and significant impact on citations, and the direction of this effect depends on the dimension of interdisciplinarity considered. Specifically, we find that a 10% increase in the Variety increases by 5.38% the number of citations received by a researcher with an article in a 5 years time period. This result is in line with Leahey et al. (2017), who finds the same positive effects on total number of citations. At the same time, we find the opposite effect for the other measures of interdisciplinarity. A 10% increase in the Balance decreases by 35.20% the citations accumulated with a paper within 5 years. This supports the idea that an even distribution of the references among fields of study negatively impacts citations, i.e. articles built on a core field of study are more easily recognized as relevant by specialized readers. Finally, a 10% increase in Disparity diminishes by 11.38% the number of citations received with an article within 5 years, suggesting that academic audiences might find it difficult to receive articles that integrate more distant knowledge, in accordance with Yegros-Yegros et al. (2015). Overall, these results suggest that IDR has a positive effect on citations if papers integrate knowledge from various, but not too distant, fields of studies while referring mainly to a specific discipline (and audience).

We continue our investigation by moving to HP1b and considering the effect of an increase in the interdisciplinary content of an article on its scientific impact when this is measured in terms of knowledge diffusion across disciplines. Estimates are conducted with the same model specifications adopted in the previous exercise, and

[18]The results are robust to controlling also for disciplines' average H-index. Results are available upon request.

Table 1.3: The effects of intedisciplinarity on citations.

| | Dependent variable: | | | |
|---|---|---|---|---|
| | log(Citations + 1) | | | |
| | (1) | (2) | (3) | (4) |
| log(Variety) | 0.647*** | 0.551*** | 0.552*** | 0.550*** |
| | (0.016) | (0.015) | (0.015) | (0.015) |
| log(Balance + 1) | −4.454*** | −4.580*** | −4.548*** | −4.554*** |
| | (0.201) | (0.191) | (0.191) | (0.191) |
| log(Disparity + 1) | −0.992*** | −1.295*** | −1.282*** | −1.268*** |
| | (0.246) | (0.229) | (0.229) | (0.229) |
| log(Number of Authors) | | 0.453*** | 0.447*** | 0.445*** |
| | | (0.013) | (0.013) | (0.013) |
| International Collaboration | | | 0.037* | 0.037* |
| | | | (0.015) | (0.015) |
| log(H-index + 1) | | | | 0.127*** |
| | | | | (0.020) |
| Variety = 1 | YES | YES | YES | YES |
| Fields of Study Fixed Effects | YES | YES | YES | YES |
| Year Fixed Effects | YES | YES | YES | YES |
| Researcher Fixed Effects | YES | YES | YES | YES |
| Number of Researchers | 6,105 | 6,105 | 6,105 | 6,105 |
| Observations | 46,156 | 46,156 | 46,156 | 46,156 |
| $R^2$ | 0.442 | 0.479 | 0.479 | 0.480 |
| Adjusted $R^2$ | 0.356 | 0.399 | 0.399 | 0.400 |

*Notes*: This table presents OLS estimates of the effects of interdisciplinarity on citations, following equation 1.1. Observations are at the paper-researcher level. The dependent variable is the logarithm of total citations accrued in five years. All regressions include individual, year and fields of study fixed effects. Standard errors are clustered at the researcher level. Significance levels: † p<0.1; * p<0.05; ** p<0.01; *** p<0.001.

the results are presented in Table 1.4.[19]

We find evidence in favor of HP1b across all model specifications: the impact of IDR on the generality of an article differs according to the interdisciplinarity

[19] We report in this table the second stage of a two-step Heckman correction model to control for potential selection in our sample (i.e. the fact that some papers have zero citations). This exercise does not rely on the use of a specific exclusion restriction, and it only makes use of those variables included in the second stage of the model (i.e. our covariates). It is worth noting that, even when an exclusion restriction is not used, identification is formally achieved, though results may be less precise in terms of statistical significance. This should be not of any practical concern, however. Our aim is to test whether our results remain qualitatively unchanged even when controlling for the potential presence of selection issues. Reassuringly, the evidence produced by our exercise confirms all our model predictions. Results of the first stage are available in Table 1.16 of the Appendix 1.8.4.

dimension considered. Specifically, we observe that an increase of Variety and Balance has a positive and statistically significant impact on the diffusion of knowledge, while the effect of Disparity is negative and statistically significant. In other words, while some dimensions of interdisciplinarity are essential for spreading ideas and concepts across multiple fields (Variety and Balance), the combinations of very distant knowledge is not well received by the scientific community.[20] All in all, we find evidence that a trade-off in public benefit exists: the diffusion of knowledge beyond disciplinary fields is favored by the increase in the degree of interdisciplinarity in scientific research, but this advantage is limited when the recombined fields of study are distant from each other.

Notably, estimates from column (4) show that even after controlling for our entire set of controls, the effect of interdisciplinarity on the diffusion of knowledge across disciplines is statistically significant, regardless of the dimension considered.[21] Moreover, the magnitude of this impact is considerable for almost all the dimensions observed. In fact, our results show that only Variety has a modest effect on generality, with a 10% increase in the number of unique fields of study in the paper's references leading to an increase of the generality index by 0.36 percentage points. On the contrary, Balance and Disparity have a sizeable positive effect on the diffusion of knowledge. In particular, a 10% increase in the Balance raise the generality index by 2.29 percentage points. On the contrary, a 10% increase in the Disparity decreases the paper generality by 1.90 percentage points.[22]

We now investigate HP2 by confronting the results from Table 1.3 and Table 1.4: i.e. comparing the direction of the effect of a given dimension of interdisciplinarity across different measures of scientific impact (number of citations vs knowledge diffusion beyond disciplinary boundaries).

We find evidence that the effect of Balance differs with the measures of scientific impact considered. In fact, papers with lower Balance have more citations but reach a less diverse audience of academics. This finding is consistent with HP2: i.e., it exists a trade-off between the number of citations that one can accrue, and the generality that one can achieve. The effect on generality corroborates what observed

[20] With respect to our previous exercise, we observe that a larger number of co-authors has a positive and statistically significant effect on the generality of the paper, while the presence of international collaborators in the team, and the H-index of the researcher, have no statistically significant effect.

[21] These results are robust to controlling also for disciplines' average H-index. Results are available upon request.

[22] As a robustness check, we have estimated the model in column (4) replacing our indicator of knowledge diffusion, i.e. the generality index, with the related and adapted Herfindahl index as developed by Gruber et al. (2013) to measure the breadth of technological recombinations in patents. We find that our results are qualitatively unchanged. Results are available upon request.

Table 1.4: The effects of intedisciplinarity on the diffusion across fields.

| | Dependent variable: log(Generality + 1) | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| log(Variety) | 0.034*** | 0.038*** | 0.038*** | 0.038*** |
| | (0.002) | (0.002) | (0.002) | (0.002) |
| log(Balance + 1) | 0.287*** | 0.237*** | 0.238*** | 0.238*** |
| | (0.024) | (0.025) | (0.025) | (0.025) |
| log(Disparity + 1) | −0.195*** | −0.201*** | −0.201*** | −0.201*** |
| | (0.038) | (0.037) | (0.037) | (0.037) |
| log(Number of Authors) | | 0.011*** | 0.011*** | 0.011*** |
| | | (0.001) | (0.001) | (0.001) |
| International Collaboration | | | 0.001 | 0.001 |
| | | | (0.001) | (0.001) |
| log(H-index + 1) | | | | −0.001 |
| | | | | (0.002) |
| IMR | −0.117*** | −0.069*** | −0.069*** | −0.069*** |
| | (0.013) | (0.015) | (0.015) | (0.015) |
| Variety = 1 | YES | YES | YES | YES |
| Fields of Study Fixed Effects | YES | YES | YES | YES |
| Year Fixed Effects | YES | YES | YES | YES |
| Researcher Fixed Effects | YES | YES | YES | YES |
| Number of Researchers | 5,938 | 5,938 | 5,938 | 5,938 |
| Observations | 44,084 | 44,084 | 44,084 | 44,084 |
| $R^2$ | 0.341 | 0.342 | 0.342 | 0.342 |
| Adjusted $R^2$ | 0.238 | 0.239 | 0.239 | 0.239 |

*Notes*: This table presents second stage results from Heckman's two-steps estimation of the effects of interdisciplinarity on the diffusion of knowledge across fields, following equation 1.1. Observations are at the paper-researcher level. The dependent variable is the logarithm of the generality index, defined in equation 1.5. All regressions include the Inverse Mills Ratio (IMR) to control for sample selection bias and individual, year and fields of study fixed effects. Standard errors are clustered at the researcher level. Significance levels: † p<0.1; * p<0.05; ** p<0.01; *** p<0.001.

for the effect of Balance on the number of citations: papers that refer more evenly to the discipline pool have no target field and, thus, for them, it is more difficult to accrue citations from a specialized literature. At the same time, these articles have a broader appeal because they bridge audiences that were previously separated, boosting the societal impact of the work.

Two relevant implications follow from our results. First, since interdisciplinarity has a statistically significant effect on citations and generality of knowledge, but the direction of the effect depends on the dimension considered, then researchers face a dilemma in how to approach IDR. In fact, despite the three dimensions are distinct, they are not completely independent. For instance, by increasing Variety (which has a positive effect on one's research impact), one will eventually increase Disparity (which has a negative effect instead).[23]

Secondly, Balance has strong but opposite effects on citations and generality. This indicates that researchers face a trade-off between increasing their reputation and reaching out to other disciplines. The costs of IDR in terms of citations are important enough to negatively impact researchers academic careers, but the public benefits regarding the diffusion of knowledge are substantive and cannot be dismissed. This disconnection between private and public returns, i.e. the interdisciplinarity dilemma, sets a challenge to the design of research policies.

### 1.6.2 Heterogeneous effects

In this section, we explore whether the effects of IDR vary according to the characteristics of the investigators, and provide different incentives to engage in interdisciplinary work. To this purpose, we estimate equation (1.1) by considering only researchers with specific features, and test whether our hypotheses (HP1a, HP1b, and HP2) are confirmed regardless of the population of scientists considered. Because of our empirical design, estimates have to be interpreted as a measure of the additional effect of an increase in the interdisciplinary content of an article on its scientific impact, given the overall effect of using an interdisciplinary content being an author with specific features: i.e., results indicate the marginal (rather than the total) effect of an increase in the interdisciplinary content of an article, given the

[23]We attempt to approach this question by including a polynomial term for the variety in our main specification and estimating their effects for both of our outcome variables. Our preliminary results show that the linear and the quadratic terms associated to variety are positive and statistically significant for both citations and generality, meaning that we do not find evidence of an optimal level of variety. Future research should be dedicated to understanding how to consider all dimensions in order to assess the optimal level of interdisciplinarity. Results are available upon request.

characteristics of the author.

We begin by testing HP1a, and results are presented in Table 1.5. In column (1) and (2), we estimate our model by considering alternatively articles written by male and female scientists. Although the effects of the IDR dimensions on the number of citations seems more pronounced for women, they are qualitatively the same. In other words, the effect of an increase in the interdisciplinary content of an article on the number of citations accrued by a female author is similar to the effect estimated for a male author: HP1a is confirmed for both of them. Of course, the fact that we find no striking differences in the effect of interdisciplinarity when separately estimating our model for women or for men, does not indicate that men and women receive the same number of citations to their articles. The total number of citations obtained by the two categories of authors for an article may still be very different. Engaging in interdisciplinary work, however, seems not to play a significant role in explaining potential differences in the scientific impact between the two categories, because all authors are subject to the same dilemma regardless of their gender.

In column (3) and (4), we estimate our model by considering alternatively articles written with or without international collaborators. This is because international collaborations may influence the heterogeneity of the team and the knowledge-integration process, which, in turn, may affect the interdisciplinarity of the article. If this is the case, it could be that these two groups are not subject in the same way to the dilemmas associated to IDR. This is not what our evidence suggests, however. In fact, we do not find any qualitative difference on the effects that the different dimensions of interdisciplinarity have on the number of citations, when separately considering these two groups. HP1a is confirmed for both of them.

In column (5) and (6), we estimate our model by considering alternatively articles written by star researchers (researchers in the upper 10th percentile of the H-index distribution within each year) and the rest of the sample. Once more, estimates are qualitatively similar between groups, and HP1a is confirmed for both of them. Prolific researchers who may engage in high-risk, high-reward publication strategies are exposed to the same effects of IDR than other researchers. Of course, this does not imply that their papers will accrue the same number of citations. This simply indicates that they face a similar dilemma, and differences across them are not to be attributed to a different effect that interdisciplinarity exerts on the scientific impact of their articles.

We continue our investigations by analyzing whether the interdisciplinary dilemma has an impact on investigators, depending on the characteristics of their co-authors in terms of IDR. In column (7) and (8), we estimate our model by considering al-

Table 1.5: Heterogeneity analysis of interdisciplinarity effects on citations.

| | Dependent variable: log(Citations + 1) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Samples | Men | Women | International Collaboration | Only US | Superstar | Non-Superstar | Interdisciplinarity Higher than Coauthors' Average | Interdisciplinarity Lower than Coauthors' Average | Coauthors with interdisciplinarity above college median | Coauthors with interdisciplinarity below college median |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| log(Variety) | 0.548*** | 0.563*** | 0.649*** | 0.532*** | 0.594*** | 0.544*** | 0.575*** | 0.493*** | 0.603*** | 0.508*** |
| | (0.017) | (0.030) | (0.041) | (0.016) | (0.044) | (0.015) | (0.020) | (0.029) | (0.027) | (0.021) |
| log(Balance + 1) | −4.541*** | −4.637*** | −4.096*** | −4.645*** | −5.328*** | −4.449*** | −5.270*** | −3.379*** | −4.673*** | −4.585*** |
| | (0.217) | (0.402) | (0.399) | (0.217) | (0.633) | (0.197) | (0.262) | (0.330) | (0.324) | (0.273) |
| log(Disparity + 1) | −1.102*** | −2.011*** | −2.921*** | −0.964*** | −1.263† | −1.256*** | −1.422*** | −1.250** | −1.797** | −1.199*** |
| | (0.252) | (0.561) | (0.593) | (0.254) | (0.760) | (0.237) | (0.358) | (0.431) | (0.583) | (0.297) |
| log(Number of Authors) | 0.444*** | 0.449*** | 0.568*** | 0.413*** | 0.527*** | 0.430*** | 0.439*** | 0.486*** | 0.431*** | 0.472*** |
| | (0.014) | (0.026) | (0.033) | (0.014) | (0.042) | (0.012) | (0.017) | (0.021) | (0.019) | (0.019) |
| International Collaboration | 0.035* | 0.039 | | | 0.049 | 0.036* | 0.037† | 0.028 | 0.063** | 0.007 |
| | (0.017) | (0.028) | | | (0.043) | (0.016) | (0.020) | (0.027) | (0.022) | (0.023) |
| log(H-index + 1) | 0.119*** | 0.136*** | 0.184** | 0.101*** | −0.171 | 0.153*** | 0.093** | 0.148*** | 0.107** | 0.122*** |
| | (0.024) | (0.034) | (0.059) | (0.021) | (0.161) | (0.020) | (0.035) | (0.035) | (0.034) | (0.032) |
| Variety = 1 | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Fields of Study Fixed Effects | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Year Fixed Effects | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Researcher Fixed Effects | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Number of Researchers | 3,926 | 2,104 | 2,599 | 5,801 | 131 | 6,074 | 3,570 | 4,080 | 3,566 | 3,886 |
| Observations | 34,614 | 11,070 | 8,761 | 37,395 | 5,137 | 41,019 | 25,832 | 15,386 | 20,224 | 20,994 |
| $R^2$ | 0.469 | 0.519 | 0.598 | 0.486 | 0.366 | 0.490 | 0.467 | 0.560 | 0.467 | 0.503 |
| Adjusted $R^2$ | 0.400 | 0.404 | 0.425 | 0.391 | 0.346 | 0.401 | 0.380 | 0.400 | 0.352 | 0.389 |

*Notes*: This table presents OLS estimates of the effects of interdisciplinarity on citations, following equation 1.1, for subsamples divided by gender (columns 1-2), the presence of international collaborators (columns 3-4), productivity (columns 5-6), coauthors interdisciplinarity (columns 7-8), and the average coauthors interdisciplinarity measured at the college level (columns 9-10). Observations are at the paper-researcher level. The dependent variable is the total citations accrued in five years. All regressions include individual, year, and fields of study fixed effects. Standard errors are clustered at the researcher level. Significance levels: †p<0.1; * p<0.05; ** p<0.01; *** p<0.001.

ternatively articles written by researchers whose interdisciplinarity profile in that year was higher or lower than the average interdisciplinarity of their UF co-authors. We define interdisciplinarity profile of researchers as the maximum value of variety registered for articles written by an investigator in one year. It may be the case that researchers face different constraints when they work with colleagues that produces papers with lower interdisciplinarity profile than when they go for a more interdisciplinary research team. However, again, estimates are qualitatively similar, which indicates that the effects of IDR on the number of citations are the same for both groups. Finally, in column (9) and (10), we estimate our model by considering alternatively articles written by researchers whose collaborators' average interdisciplinarity in a year was higher or lower than the average interdisciplinary of the college in which they are affiliated. Overall, we do not find evidence that the effects of IDR are driven by co-authors interdisciplinarity[24].

We further proceed by estimating equation (1.1) when considering a sample of researchers affiliated to a specific academic unit, in which the researcher's college and department are included. Results are reported in Table 1.6. In the estimations presented in this table, we alternatively consider researchers affiliated to: the College of Liberal Arts and Science (CLAS), column (1); the College of Engineering (ENG), column (2); the Health Science Center (HSC), column (3); and the Institute of Food and Agricultural Sciences (IFAS).[25] Results are qualitatively unchanged regardless of the affiliation considered, hinting that IDR has the same effect on the number of citations in all academic environments: i.e., researchers are subject to the same dilemma regardless of their affiliation, and HP1a is confirmed for all of them.[26]

We now replicate our exercise by measuring scientific impact in terms of the diffusion of knowledge across disciplines, to test HP1b. Our results are reported in Table 1.7 and Table 1.8. Also in this case, we find evidence that researchers face the same dilemma when engaging in interdisciplinary work, regardless of their specific characteristics[27]. At the same time, when considering researchers' affiliation,

[24]As a robustness check, we have estimated the model in columns (7-9) in Table 1.5 adding to our specification a dummy variable registering if a paper was co-authored exclusively by UF investigators and interactions between this dummy and our interdisciplinarity measures. We find that our results are qualitatively unchanged. Results are available upon request.

[25]The colleges included in each academic unit are reported in Table 1.9 in the Appendix 3.7.1.

[26]We also estimate equation (1.1) using alternative disciplinary subdivisions based on researchers' paper fields of study, individual main field of publication (measured as the field where the researcher published most of her papers), and also using department-level affiliation. All our results are qualitatively unchanged. Results are available upon request.

[27]As a robustness check, we have estimated the model in columns (7-9) in Table 1.7 adding to our specification a dummy variable registering if a paper was co-authored exclusively by UF investigators and interactions between this dummy and our interdisciplinarity measures. We find that our results are qualitatively unchanged. Results are available upon request.

Table 1.6: Interdisciplinarity effects on citations and college affiliation.

| | Dependent variable: log(Citations + 1) | | | |
|---|---|---|---|---|
| | CLAS | ENG | HSC | IFAS |
| | (1) | (2) | (3) | (4) |
| log(Variety) | 0.489*** | 0.493*** | 0.567*** | 0.555*** |
| | (0.048) | (0.046) | (0.024) | (0.030) |
| log(Balance + 1) | −5.028*** | −3.781*** | −4.930*** | −3.631*** |
| | (0.572) | (0.770) | (0.283) | (0.373) |
| log(Disparity + 1) | −1.975** | −1.456* | −1.089** | −1.457* |
| | (0.636) | (0.563) | (0.337) | (0.658) |
| log(Number of Authors) | 0.449*** | 0.317*** | 0.506*** | 0.343*** |
| | (0.031) | (0.058) | (0.018) | (0.025) |
| International Collaboration | −0.048 | 0.023 | 0.070** | 0.092** |
| | (0.040) | (0.041) | (0.026) | (0.030) |
| log(H-index + 1) | 0.066 | 0.084 | 0.164*** | 0.040 |
| | (0.069) | (0.073) | (0.030) | (0.035) |
| Variety = 1 | YES | YES | YES | YES |
| Fields of Study Fixed Effects | YES | YES | YES | YES |
| Year Fixed Effects | YES | YES | YES | YES |
| Researcher Fixed Effects | YES | YES | YES | YES |
| Number of Researchers | 665 | 389 | 2,215 | 1,198 |
| Observations | 5,267 | 4,946 | 18,260 | 9,825 |
| $R^2$ | 0.503 | 0.360 | 0.474 | 0.478 |
| Adjusted $R^2$ | 0.427 | 0.301 | 0.401 | 0.403 |

*Notes*: This table presents OLS estimates of the effects of interdisciplinarity on citations, following equation 1.1, for subsamples divided by academic unit affiliation. Column 1 estimates the effects for researchers affiliated to the College of Liberal Arts and Science, column 2 for those in the College of Engineering, column 3 for those on the Health Science Center and column 4 for those in the Institute of Food and Agricultural Sciences. Observations are at the paper-researcher level. The dependent variable is the total citations accrued in five years. All regressions include individual, year and fields of study fixed effects. Standard errors are clustered at the researcher level. Significance levels: † p<0.1; * p<0.05; ** p<0.01; *** p<0.001.

we observe that the effect of Disparity is negative but no longer significant for the researchers in the College of Liberal Arts and Science (column (1)). This may suggest that researchers in social sciences, humanities, and hard sciences like physics are not penalized as much for combining dissimilar disciplines as more "applied" fields like engineering, health, and agricultural science.[28] In all the other cases, however,

[28]Differences might be due to evaluation criteria that vary across sciences. For instance, Guetzkow et al. (2004) maintain that social sciences and humanities rely mainly on originality that, in

Table 1.7: Heterogeneity analysis of interdisciplinarity effects on diffusion across fields.

| Samples | Dependent variable: log(Generality + 1) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Men | Women | International Collaboration | Only US | Superstar | Non-Superstar | Interdisciplinarity Higher than Coauthors' Average | Interdisciplinarity Lower than Coauthors' Average | Coauthors with interdisciplinarity above college median | Coauthors with interdisciplinarity below college median |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| log(Variety) | 0.037*** | 0.045*** | 0.028*** | 0.041*** | 0.038*** | 0.040*** | 0.040*** | 0.035*** | 0.037*** | 0.040*** |
| | (0.003) | (0.005) | (0.005) | (0.003) | (0.005) | (0.003) | (0.003) | (0.005) | (0.003) | (0.004) |
| log(Balance + 1) | 0.244*** | 0.221*** | 0.309*** | 0.217*** | 0.260** | 0.228*** | 0.265*** | 0.236*** | 0.280*** | 0.220*** |
| | (0.029) | (0.047) | (0.042) | (0.029) | (0.080) | (0.026) | (0.034) | (0.046) | (0.035) | (0.040) |
| log(Disparity + 1) | −0.149*** | −0.432*** | −0.147 | −0.195*** | −0.280* | −0.198*** | −0.207*** | −0.230*** | −0.215** | −0.231*** |
| | (0.042) | (0.078) | (0.092) | (0.042) | (0.109) | (0.040) | (0.056) | (0.064) | (0.078) | (0.051) |
| log(Number of Authors) | 0.011*** | 0.012*** | 0.008*** | 0.013*** | 0.013*** | 0.011*** | 0.012*** | 0.012*** | 0.014*** | 0.011*** |
| | (0.002) | (0.003) | (0.002) | (0.002) | (0.003) | (0.001) | (0.002) | (0.003) | (0.002) | (0.002) |
| International Collaboration | 0.0003 | 0.003 | | | 0.001 | 0.001 | −0.0003 | 0.001 | 0.001 | 0.001 |
| | (0.002) | (0.003) | | | (0.003) | (0.002) | (0.002) | (0.003) | (0.002) | (0.002) |
| log(H-index + 1) | −0.004 | 0.004 | −0.009* | −0.001 | −0.021 | −0.001 | −0.007† | 0.005 | −0.006 | 0.001 |
| | (0.003) | (0.004) | (0.005) | (0.003) | (0.014) | (0.002) | (0.004) | (0.005) | (0.004) | (0.004) |
| IMR | −0.073*** | −0.064† | −0.165*** | −0.048** | −0.003 | −0.067*** | −0.070*** | −0.070* | −0.074** | −0.051* |
| | (0.016) | (0.033) | (0.048) | (0.017) | (0.040) | (0.016) | (0.021) | (0.028) | (0.025) | (0.023) |
| Variety = 1 | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Fields of Study Fixed Effects | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Year Fixed Effects | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Researcher Fixed Effects | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Number of Researchers | 3,836 | 2,028 | 2,562 | 5,631 | 131 | 5,907 | 3,499 | 3,994 | 3,533 | 3,769 |
| Observations | 33,007 | 10,618 | 8,561 | 35,523 | 4,974 | 39,110 | 24,806 | 14,737 | 19,684 | 19,859 |
| $R^2$ | 0.326 | 0.397 | 0.558 | 0.347 | 0.206 | 0.355 | 0.332 | 0.457 | 0.372 | 0.364 |
| Adjusted $R^2$ | 0.236 | 0.252 | 0.366 | 0.223 | 0.180 | 0.240 | 0.221 | 0.253 | 0.233 | 0.213 |

*Notes*: This table presents second stage results from Heckman's two-steps estimation of the effects of interdisciplinarity on the diffusion of knowledge across fields, following equation 1.1, for subsamples divided by gender (columns 1-2), the presence of international collaborators (columns 3-4), productivity (columns 5-6), coauthors interdisciplinarity (columns 7-8), and the average coauthors interdisciplinarity measured at the college level (columns 9-10). Observations are at the paper-researcher level. The dependent variable is the logarithm of the generality index, defined in equation 1.5. All regressions include the Inverse Mills Ratio (IMR) to control for sample selection bias and individual, year, and fields of study fixed effects. Standard errors are clustered at the researcher level. Significance levels: †p<0.1; * p<0.05; ** p<0.01; *** p<0.001.

Table 1.8: Interdisciplinarity effects on diffusion across fields and college affiliation.

| | Dependent variable: log(Generality + 1) | | | |
|---|---|---|---|---|
| | CLAS | ENG | HSC | IFAS |
| | (1) | (2) | (3) | (4) |
| log(Variety) | 0.026*** | 0.037*** | 0.037*** | 0.051*** |
| | (0.005) | (0.008) | (0.004) | (0.006) |
| log(Balance + 1) | 0.099* | 0.275** | 0.269*** | 0.214*** |
| | (0.048) | (0.100) | (0.040) | (0.058) |
| log(Disparity + 1) | −0.101 | −0.187† | −0.140* | −0.463*** |
| | (0.070) | (0.100) | (0.063) | (0.105) |
| log(Number of Authors) | 0.014*** | 0.014* | 0.010*** | 0.006† |
| | (0.003) | (0.006) | (0.002) | (0.003) |
| International Collaboration | −0.009* | 0.001 | 0.005** | 0.005 |
| | (0.004) | (0.005) | (0.002) | (0.003) |
| log(H-index + 1) | −0.007 | 0.004 | 0.006 | −0.015** |
| | (0.007) | (0.008) | (0.003) | (0.005) |
| IMR | −0.038 | −0.020 | −0.088*** | −0.065† |
| | (0.031) | (0.055) | (0.022) | (0.039) |
| Variety = 1 | YES | YES | YES | YES |
| Fields of Study Fixed Effects | YES | YES | YES | YES |
| Year Fixed Effects | YES | YES | YES | YES |
| Researcher Fixed Effects | YES | YES | YES | YES |
| Number of Researchers | 620 | 385 | 2,151 | 1,172 |
| Observations | 5,024 | 4,741 | 17,389 | 9,368 |
| $R^2$ | 0.386 | 0.243 | 0.285 | 0.335 |
| Adjusted $R^2$ | 0.294 | 0.171 | 0.183 | 0.237 |

*Notes*: This table presents second stage results from Heckman's two-steps estimation of the effects of interdisciplinarity on the diffusion of knowledge across fields, 1.1, for subsamples divided by academic unit affiliation. Column 1 estimates the effects for researchers affiliated to the College of Liberal Arts and Science, column 2 for those in the College of Engineering, column 3 for those on the Health Science Center and column 4 for those in the Institute of Food and Agricultural Sciences. Observations are at the paper-researcher level. The dependent variable is the total citations accrued in five years. All regressions include the Inverse Mills Ratio (IMR) to control for sample selection bias and individual, year and fields of study fixed effects. Standard errors are clustered at the researcher level. Significance levels: † $p<0.1$; * $p<0.05$; ** $p<0.01$; *** $p<0.001$.

we still find evidence that Variety and Balance have a positive and statistically significant impact on the diffusion of knowledge across disciplines for all considered subgroups, while Disparity has a negative and statistically significant effect.

Finally, we investigate the validity of HP2 by comparing the results obtained

their study, includes disciplinary variation.

when using different measures of scientific impact. Notably, we still find that the Balance has strong but opposite effects on citations and generality of knowledge. Papers with lower Balance have more citations but reach a less diverse audience of academics, regardless of the characteristics of the group considered. In other words, we find evidence that HP2 applies to all research profiles: i.e., all of them are subject to an interdisciplinary dilemma in their work.

Taken together, our results suggest that all scholars face the similar incentives and constraints in engaging in more interdisciplinary projects. Regardless of their characteristics or affiliation, the effects of IDR are large and widespread, and affect all research activities at the University of Florida.

## 1.7 Conclusion

Our results bring evidence to the idea that multiple logics within the academia might create contrasting incentives for scholars. In our study, we highlight that policies that govern hiring and evaluation within universities and policy that sustain interdisciplinarity incentivize behaviors that are, at least to a certain extent, incompatible. It is not always possible to act as to accumulate citations from published papers while combining knowledge from different domains: scholars are forced to trade-off between reputation and societal impact of their research.

Nowadays, the soaring amount of knowledge accumulated in published articles requires doctoral programs and post-doctoral training of longer duration (Jones, 2010). This, in turn, postpone first publishing (Conti and Liu, 2015) and the 'age of great achievement' (Jones et al., 2014; The National Academies, 1998). To compensate for such burden of knowledge, scientists often seek to (over)specialize in specific fields, making interdisciplinarity a necessary choice to ensure scientific communication and societal progress. Therefore, the need for appropriate incentives and coordinated policy is particularly urgent.

As the literature on IDR grows and gets more sophisticated, we encourage more investigations at the level of researchers to fully grasp the implications of choosing an interdisciplinary approach. Although our results are robust to various specifications and definitions of scholars' samples, further research is needed to corroborate the external validity of our analysis by including information on research activities in more than one university.

## Acknowledgements

This research uses data from the Bureau of Economics and Business Research (BEBR) of the University of Florida. We thank the staff of the BEBR and the director Christopher McCarty for their important contribution to this work. We are also grateful for helpful comments from seminar participants at the Regional Quantitative Analysis Group (AQR) seminar series of the University of Barcelona.

## 1.8 Appendix for Chapter 1

In this appendix, we report the procedure we followed to construct the database used in the analysis and some additional descriptive statistics.

### 1.8.1 Data collection

From the data collected by the Bureau of Economics and Business Research (BEBR) of the University of Florida (UF), we retrieved information concerning publication records, department affiliation, and gender for the universe of UF's researchers in the period 2008-2013. Each researcher is identified by a unique code (UFID). Raw publication records provide information regarding 34,851 scholarly works including journal name, article title, and publication year. Based on UF registered publications' information, we retrieved the publications' Digital Object Identifiers (DOI) from Crossref and Scopus databases.[29] This procedure allows us to identify researchers' academic output that was indexed in the largest and most common scholarly works' databases.

More specifically, we used an automated script to extract bibliographic metadata of UF publications available in the original dataset through Scopus Database API Interface and Crossref REST API.[30] The three main steps of this procedure are the following:

1 Get articles partial metadata based on publication title: From titles of publications in the UF records, the script – through queries to Scopus and Crossref APIs – collects publications matching our list of articles' titles and retrieves their metadata (DOI, journal name, publication title, publication year). We collect the first ten results of the queries for each title and store them in a new database.

2 Cleaning and processing article's title: The article titles in the raw data and in the data retrieved by API queries are cleaned and then processed. Cleaning consists in eliminating spaces, special characters, and punctuation. Processing consists in coercing characters to lowercase and comparing the raw (original) and newly extracted titles.

3 Title-DOI matching procedure: Matches are determined according to a fuzzy matching algorithm implemented in the *fuzzywuzzy* text similarity package

[29] The databases are available at the following webpage: Crossref and Scopus.

[30] Data collection using Scopus and Crossref occurred in 2018.

in Python.[31] The script considers a match if titles have a higher than 90% similarity ratio and the matching is unique. Matched publications and its respective metadata are assigned to the associated researcher. Unique matches with more than one DOI were manually checked and disambiguated. Publications without a unique match are dropped.

With this procedure, we were able to identify the DOIs of 28,239 publications of our original database. Using these DOIs, we collect the full metadata through Lens and Microsoft Academic Graph (MAG) databases.[32] Metadata from Lens API platform includes: IDs (Lens articles ID, Microsoft Academic Graph ID); publication type (journal article, book, working paper); list of citations; list of references; fields of study (computed by the MAG algorithm as described in Section 1.5); and authors' affiliations. We decided to focus on Lens database to collect citations and references data because it also provides their disciplines based on the natural language processing algorithms used by MAG. Furthermore, the Lens' scholarly citation data, contrary to Microsoft Academic Graph, indexes only publications of selected document types (journal article, book, working paper).[33] Publications missing references or missing fields of study are dropped. In addition, we restrict our sample to only journal articles. Our final database consists of 23,926 articles and their full metadata.

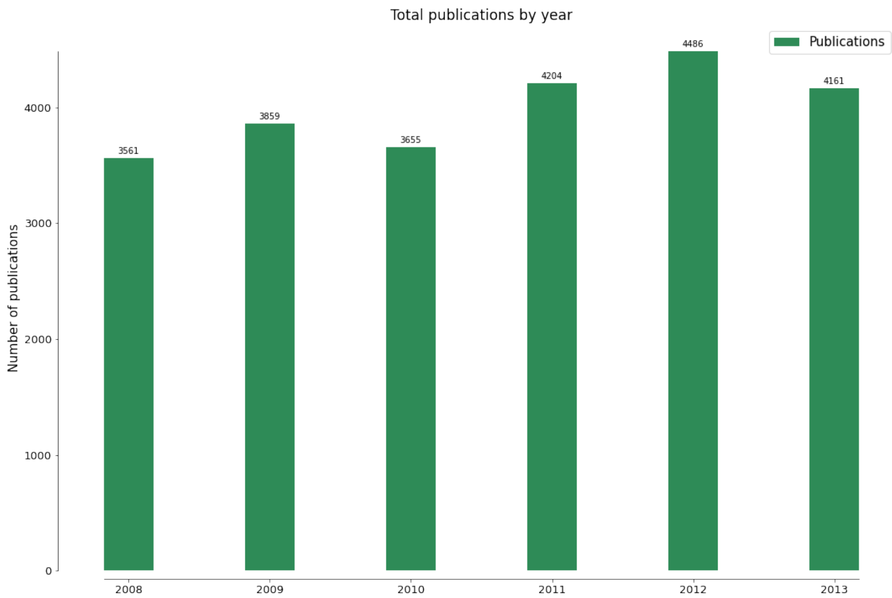
In the last data collection phase, we extracted from MAG a proximity measure between the fields of study using the functionality Network Similarity Package, as described in Section 1.5.[34] We collected similarity scores for all possible combinations between the 19 fields (first level of classification) and 292 subfields (second level of classification).

### 1.8.2 Additional descriptive statistics

Figure 1.5 shows the evolution of the total number of publications in our database (in the period 2008–2013). Table 1.9 reports the distribution of researchers across academic units and colleges, while 1.10 shows the distribution of citations by field of study. Table 1.11 shows summary statistics at the paper-researcher level. Finally, Table 1.12 reports the correlation between variables used in our regression analysis.

[31] Documentation about *fuzzywuzzy* is available here: fuzzywuzzy.
[32] These databases are available at the following link: Microsoft Academic Graph and Lens.
[33] Data collection using the Lens occurred in 2019.
[34] Data collection Microsoft Academic Graph occurred in 2020.

Figure 1.5: Number of total publications by year.




### 1.8.3 Motivating evidence

This section discusses data about wages and grants received by a sample of 3,481 UF's researchers. We exploit this data to conduct a preliminary investigation on researchers' trade-offs in pursing IDR (see Section 1.2).

To perform this analysis, we compute an aggregated indicator of interdisciplinarity profile at the researcher level, since the yearly wage and the number of grants refer to scholars. This indicator is equal to the maximum value of the number of cited fields of study (second level of MAG classification) found among the articles written by a researcher in a given year. We rely on MAG, instead, to define an indicator of scholar seniority: the variable academic age measures the time that a researcher has been active and is defined as the number of years between their first published work until the year of observation. Table 1.13 shows descriptive statistics for these variables, while Table 1.14 reports the results of our preliminary regression analysis.

Table 1.9: Number of researchers by academic units and colleges.

| **Academic Units** | **Colleges** | **Researchers** |
|---|---|---|
| Liberal Arts and Sciences | College of Liberal Arts and Sciences | 665 |
| Engineering | College of Engineering | 389 |
| Health Sciences | Medicine | 1545 |
| | Medicine-Jacksonville | 175 |
| | Public Health and Health Professions | 166 |
| | Pharmacy | 140 |
| | Dentistry | 139 |
| | Nursing | 36 |
| | Health Affairs | 14 |
| Food and Agricultural Sciences | Agricultural And Life Sciences | 978 |
| | Veterinary Medicine | 218 |
| | Institute of Food and Agricultural Sciences | 2 |
| Other | UF Students | 946 |
| | Uncategorized Departments | 687 |

*Notes*: This table shows the distribution researchers affiliated to the academic units and colleges at the University of Florida (UF) from 2008 to 2013. The total number of researchers with a college affiliation is 5130. Researchers that are not affiliated to any specific academic unit are counted in the category "Other". Researchers classified as students in the UF registry office are counted in "UF Students" and faculty affiliated to departments not belonging to any college are counted in "Uncategorized Departments".

Table 1.10: Distribution of citations by field of study (first level of classification).

| **Field of Study** | **References** | **Citations** |
|---|---|---|
| Art | 5490 | 1393 |
| Biology | 1181592 | 359734 |
| Business | 21831 | 6804 |
| Chemistry | 329228 | 101792 |
| Computer science | 88408 | 23706 |
| Economics | 64521 | 15417 |
| Engineering | 79673 | 28339 |
| Environmental science | 40883 | 14674 |
| Geography | 27317 | 6791 |
| Geology | 89619 | 25773 |
| History | 8581 | 1026 |
| Materials science | 54980 | 29450 |
| Mathematics | 101585 | 19532 |
| Medicine | 1071812 | 378978 |
| Philosophy | 10873 | 2082 |
| Physics | 225173 | 78280 |
| Political science | 10374 | 2585 |
| Psychology | 233354 | 62381 |
| Sociology | 38379 | 7401 |

*Notes*: This table shows the distribution of the documents of each field in the focal papers' references and which cited our focal paper (citations). The total number of documents referenced is 646,280 and the total number of citations is 366,024

Table 1.11: Summary statistics at the article-researcher level.

| **Variables** | *Mean* | *SD* | *Min* | *50%* | *Max* | *Obs* |
|---|---|---|---|---|---|---|
| Nb. Citations | 20.73 | 43.02 | 0 | 11 | 2,530 | 46,156 |
| Generality | 0.73 | 0.17 | 0 | 0.78 | 0.98 | 44,084 |
| Variety | 39.56 | 18.97 | 1 | 39 | 153 | 46,156 |
| Balance | 0.84 | 0.08 | 0 | 0.85 | 1 | 46,156 |
| Disparity | 0.69 | 0.06 | 0 | 0.71 | 0.94 | 46,156 |
| Nb. of Authors | 6.30 | 7.98 | 1 | 5 | 1,269 | 46,156 |
| International Collab. | 0.19 | 0.39 | 0 | 0 | 1 | 46,156 |
| H-index | 6.29 | 7.00 | 0 | 4 | 54 | 46,156 |

*Notes*: Nb. Citations is the total number of citations received in a 5 years time after the publication. The Generality captures the degree of applicability of the knowledge codified in a paper on different fields of study. It is worth noting that generality is not defined for papers with zero citations. International collaboration is a dummy variable that assumes the value 1 when at least one co-author in the paper is affiliated to an institution outside the United States.

Table 1.12: Correlation table between variables used in regressions.

| | Number Citations | Generality | Variety | Balance | Disparity | Number References | Number Authors | International Collaboration |
|---|---|---|---|---|---|---|---|---|
| Number Citations | 1 | 0.18 | 0.18 | -0.12 | 0.09 | 0.33 | 0.45 | 0.11 |
| Generality | 0.18 | 1 | 0.23 | 0.12 | 0.12 | 0.16 | 0.06 | 0.03 |
| Variety | 0.18 | 0.23 | 1 | -0.05 | 0.57 | 0.67 | 0.07 | 0.04 |
| Balance | -0.12 | 0.12 | -0.05 | 1 | 0.10 | -0.41 | -0.08 | -0.11 |
| Disparity | 0.09 | 0.12 | 0.57 | 0.10 | 1 | 0.28 | 0.06 | 0.02 |
| Number References | 0.33 | 0.16 | 0.67 | -0.41 | 0.28 | 1 | 0.20 | 0.09 |
| Number of Authors | 0.45 | 0.06 | 0.07 | -0.08 | 0.06 | 0.20 | 1 | 0.16 |
| International Collaboration | 0.11 | 0.03 | 0.04 | -0.11 | 0.02 | 0.09 | 0.16 | 1 |

Table 1.13: Additional descriptive statistics on wages and grants.

| **Variables** | *Mean* | *SD* | *Min* | *50%* | *Max* | *Obs* |
|---|---|---|---|---|---|---|
| Wage | 127199.10 | 90411.25 | 100 | 99553.2 | 997465 | 11,160 |
| Nb. Grants | 0.55 | 1.25 | 0 | 0 | 21 | 18,005 |
| Interdisciplinarity | 46.34 | 20.69 | 1 | 46 | 153 | 18,005 |
| Academic Age | 15.54 | 11.44 | 1 | 13 | 53 | 14,606 |

*Notes*: These variable are available only for a subsample of UF's researchers. The variable interdisciplinarity is equal to the maximum value of the number of cited fields of study found among the articles written by a researcher in a given year. Academic age measures the time that a researcher has been active in a research field and is defined as the number of years between their first published work until the year of observation.

Table 1.14: Correlation between scholars' interdisciplinarity profile and academic achievements.

| | Dependent variable: | | | |
|---|---|---|---|---|
| | log(Wage) | | log(Nb. Grants + 1) | |
| | (1) | (2) | (3) | (4) |
| Interdisciplinarity | $-0.001^{*}$ | $-0.002^{***}$ | $0.004^{***}$ | $0.003^{***}$ |
| | (0.0003) | (0.0003) | (0.0003) | (0.0003) |
| Academic Age | | $0.018^{***}$ | | $0.010^{***}$ |
| | | (0.001) | | (0.001) |
| Constant | $11.578^{***}$ | $11.309^{***}$ | $0.224^{***}$ | $0.118^{***}$ |
| | (0.015) | (0.018) | (0.013) | (0.017) |
| Number of Researchers | 3,481 | 2,785 | 3,481 | 2,785 |
| Observations | 11,160 | 9,444 | 11,160 | 9,444 |
| $R^2$ | 0.0004 | 0.091 | 0.022 | 0.045 |
| Adjusted $R^2$ | 0.0003 | 0.091 | 0.022 | 0.045 |

*Notes*: estimated coefficients and standard errors (parentheses) obtained with ordinary least square estimations. The dependent variables are the logarithm of yearly wages (columns 1-2) and the number of awarded grants to a researcher in a year (columns 3-4). The variable interdisciplinarity is equal to the maximum value of the number of cited fields of study found among the articles written by a researcher in a given year. Academic age measures the time that a researcher has been active and is defined as the number of years between their first published work until the year of observation. Significance levels: † $p<0.1$; * $p<0.05$; ** $p<0.01$; *** $p<0.001$.

### 1.8.4 Robustness checks and first stage of Heckman correction

Table 1.15 allows comparing the estimations obtained through the use of OLS with those resulting from Poisson and negative binomial. As evident in the table, our results are robust to the different estimation approaches.

Table 1.15: Results using OLS, Poisson, and negative binomial to estimate the effect of IDR con the number of citations.

| Dependent Variables: | log(Nb. of Citations+1) | Nb. of Citations | Nb. of Citations |
|---|---|---|---|
| | (1) | (2) | (3) |
| | *OLS* | *Poisson* | *Neg. Bin.* |
| log(Variety) | 0.5499*** | 0.6920*** | 0.6211*** |
| | (0.0146) | (0.0350) | (0.0226) |
| log(Balance + 1) | -4.552*** | -4.410*** | -4.526*** |
| | (0.1910) | (0.4070) | (0.2432) |
| log(Disparity + 1) | -1.268*** | -2.296*** | -1.518*** |
| | (0.2287) | (0.5492) | (0.3545) |
| log(Number of Authors) | 0.4454*** | 0.5343*** | 0.4793*** |
| | (0.0126) | (0.0351) | (0.0192) |
| International Collaboration | 0.0376** | 0.1133*** | 0.0615*** |
| | (0.0149) | (0.0360) | (0.0227) |
| log(H-index + 1) | 0.1266*** | 0.2715*** | 0.1506*** |
| | (0.0196) | (0.0328) | (0.0233) |
| Variety = 1 | YES | YES | YES |
| Fields of Study Fixed Effects | YES | YES | YES |
| Year Fixed Effects | YES | YES | YES |
| Researcher Fixed Effects | YES | YES | YES |
| Observations | 46,159 | 45,974 | 45,974 |
| Squared Correlation | 0.47950 | 0.39767 | 0.31271 |
| Pseudo $R^2$ | 0.21431 | 0.44458 | 0.09057 |
| BIC | 176,539.3 | 954,572.6 | 402,511.9 |
| Over-dispersion | | | 1.6172 |

Note: S.E. clustered at researcher level *p<0.1; **p<0.05; ***p<0.01

Table 1.16, instead, reports the first stage of Heckman correction used to estimate the effect of IDR on Generality. Generality is, indeed, only defined for articles that receive at least one citation.

Table 1.16: First stage of the Heckman correction.

| | Dependent variable: | | | |
|---|---|---|---|---|
| | Cited Paper | | | |
| | *probit* | | | |
| | (1) | (2) | (3) | (4) |
| log(Variety) | 0.762*** | 0.653*** | 0.653*** | 0.645*** |
| | (0.023) | (0.023) | (0.023) | (0.023) |
| log(Balance + 1) | −6.193*** | −6.054*** | −6.023*** | −5.916*** |
| | (0.327) | (0.331) | (0.332) | (0.332) |
| log(Disparity + 1) | −1.479*** | −1.797*** | −1.782*** | −1.769*** |
| | (0.315) | (0.324) | (0.324) | (0.324) |
| log(Number of Authors) | | 0.496*** | 0.491*** | 0.479*** |
| | | (0.021) | (0.021) | (0.021) |
| International Collaboration | | | 0.055 | 0.040 |
| | | | (0.037) | (0.037) |
| log(H-index + 1) | | | | 0.115*** |
| | | | | (0.016) |
| Constant | 3.836*** | 3.571*** | 3.545*** | 3.424*** |
| | (0.238) | (0.243) | (0.243) | (0.243) |
| Variety = 1 | YES | YES | YES | YES |
| Fields of Study Fixed Effects | YES | YES | YES | YES |
| Year Fixed Effects | YES | YES | YES | YES |
| Researcher Fixed Effects | NO | NO | NO | NO |
| Number of Researchers | 6,105 | 6,105 | 6,105 | 6,105 |
| Observations | 46,173 | 46,156 | 46,156 | 46,156 |
| Log Likelihood | -6,555.652 | -6,247.971 | -6,246.868 | -6,219.912 |
| Akaike Inf. Crit. | 13,195.300 | 12,581.940 | 12,581.740 | 12,529.830 |

*Notes*: This table presents first stage results from Heckman's two-steps estimation. Observations are at the paper-researcher level. Estimates stem from probit specifications with dependent variable being a dummy that assumes value 1 when a paper was cited in a 5 year time-window after the publication and 0 otherwise. All regressions include year and fields of study fixed effects. Standard errors are clustered at the researcher level. Significance levels: † p<0.1; * p<0.05; ** p<0.01; *** p<0.001.

### 1.8.5 Fields of study classification

Table 1.17 reports the conversion table, made by authors, between the first and the second level of fields of studies, as classified by MAG. The first level classify articles in 19 disciplines, while the second one has 292 possible values, corresponding to sub-disciplines. The table also include the ID used to represent fields of studies at the second level in the knowledge space (Figure 1.3).

| ID | $2^{nd}$ level | $1^{st}$ level |
|---|---|---|
| 0 | Visual arts | Art |
| 1 | Classics | Art |
| 2 | Art history | Art |
| 3 | Literature | Art |
| 4 | Linguistics | Art |
| 5 | Communication | Art |
| 6 | Library science | Art |
| 7 | Humanities | Art |
| 8 | Zoology | Biology |
| 9 | Botany | Biology |
| 10 | Evolutionary biology | Biology |
| 11 | Computational biology | Biology |
| 12 | Cell biology | Biology |
| 13 | Molecular biology | Biology |
| 14 | Animal science | Biology |
| 15 | Astrobiology | Biology |
| 16 | Microbiology | Biology |
| 17 | Food science | Biology |
| 18 | Biotechnology | Biology |
| 19 | Biological system | Biology |
| 20 | Economic system | Business |
| 21 | Financial system | Business |
| 22 | Commerce | Business |
| 23 | Knowledge management | Business |
| 24 | Process management | Business |
| 25 | Marketing | Business |
| 26 | Public relations | Business |
| 27 | Advertising | Business |
| 28 | Accounting | Business |
| 29 | Operations research | Business |
| 30 | Management | Business |
| 31 | Operations management | Business |
| 32 | Management science | Business |
| 33 | Business administration | Business |
| 34 | Geochemistry | Chemistry |

| ID | $2^{nd}$ level | $1^{st}$ level |
|---|---|---|
| 35 | Computational chemistry | Chemistry |
| 36 | Physical chemistry | Chemistry |
| 37 | Organic chemistry | Chemistry |
| 38 | Stereochemistry | Chemistry |
| 39 | Environmental chemistry | Chemistry |
| 40 | Inorganic chemistry | Chemistry |
| 41 | Photochemistry | Chemistry |
| 42 | Combinatorial chemistry | Chemistry |
| 43 | Polymer chemistry | Chemistry |
| 44 | Analytical chemistry | Chemistry |
| 45 | Medicinal chemistry | Chemistry |
| 46 | Biochemistry | Chemistry |
| 47 | Nuclear chemistry | Chemistry |
| 48 | Chromatography | Chemistry |
| 49 | Radiochemistry | Chemistry |
| 50 | Toxicology | Chemistry |
| 51 | Pharmacology | Chemistry |
| 52 | Embedded system | Computer science |
| 53 | Distributed computing | Computer science |
| 54 | Computer network | Computer science |
| 55 | Artificial intelligence | Computer science |
| 56 | Pattern recognition | Computer science |
| 57 | Computer vision | Computer science |
| 58 | Machine learning | Computer science |
| 59 | Real-time computing | Computer science |
| 60 | World Wide Web | Computer science |
| 61 | Information retrieval | Computer science |
| 62 | Internet privacy | Computer science |
| 63 | Computer security | Computer science |
| 64 | Operating system | Computer science |
| 65 | Human–computer interaction | Computer science |
| 66 | Multimedia | Computer science |
| 67 | Natural language processing | Computer science |
| 68 | Data mining | Computer science |
| 69 | Programming language | Computer science |
| 70 | Theoretical computer science | Computer science |
| 71 | Algorithm | Computer science |
| 72 | Data science | Computer science |
| 73 | Database | Computer science |
| 74 | Bioinformatics | Computer science |
| 75 | Parallel computing | Computer science |
| 76 | Computer graphics (images) | Computer science |
| 77 | Computational science | Computer science |
| 78 | Speech recognition | Computer science |
| 79 | International economics | Economics |

| ID | $2^{nd}$ level | $1^{st}$ level |
|---|---|---|
| 80 | International trade | Economics |
| 81 | Market economy | Economics |
| 82 | Econometrics | Economics |
| 83 | Macroeconomics | Economics |
| 84 | Monetary economics | Economics |
| 85 | Economic policy | Economics |
| 86 | Positive economics | Economics |
| 87 | Neoclassical economics | Economics |
| 88 | Industrial organization | Economics |
| 89 | Finance | Economics |
| 90 | Natural resource economics | Economics |
| 91 | Environmental economics | Economics |
| 92 | Keynesian economics | Economics |
| 93 | Political economy | Economics |
| 94 | Development economics | Economics |
| 95 | Economic history | Economics |
| 96 | Agricultural economics | Economics |
| 97 | Economy | Economics |
| 98 | Financial economics | Economics |
| 99 | Labour economics | Economics |
| 100 | Demographic economics | Economics |
| 101 | Law and economics | Economics |
| 102 | Economic growth | Economics |
| 103 | Public economics | Economics |
| 104 | Microeconomics | Economics |
| 105 | Classical economics | Economics |
| 106 | Mathematical economics | Economics |
| 107 | Welfare economics | Economics |
| 108 | Computer hardware | Engineering |
| 109 | Electronic engineering | Engineering |
| 110 | Electrical engineering | Engineering |
| 111 | Systems engineering | Engineering |
| 112 | Software engineering | Engineering |
| 113 | Control engineering | Engineering |
| 114 | Control theory | Engineering |
| 115 | Environmental engineering | Engineering |
| 116 | Mechanics | Engineering |
| 117 | Manufacturing engineering | Engineering |
| 118 | Industrial engineering | Engineering |
| 119 | Mechanical engineering | Engineering |
| 120 | Engineering drawing | Engineering |
| 121 | Aerospace engineering | Engineering |
| 122 | Aeronautics | Engineering |
| 123 | Construction engineering | Engineering |
| 124 | Engineering management | Engineering |

| ID | $2^{nd}$ level | $1^{st}$ level |
|---|---|---|
| 125 | Geotechnical engineering | Engineering |
| 126 | Civil engineering | Engineering |
| 127 | Pulp and paper industry | Engineering |
| 128 | Structural engineering | Engineering |
| 129 | Agricultural engineering | Engineering |
| 130 | Optoelectronics | Engineering |
| 131 | Computer architecture | Engineering |
| 132 | Architectural engineering | Engineering |
| 133 | Chemical engineering | Engineering |
| 134 | Risk analysis (engineering) | Engineering |
| 135 | Reliability engineering | Engineering |
| 136 | Computer engineering | Engineering |
| 137 | Transport engineering | Engineering |
| 138 | Process engineering | Engineering |
| 139 | Biochemical engineering | Engineering |
| 140 | Petroleum engineering | Engineering |
| 141 | Automotive engineering | Engineering |
| 142 | Telecommunications | Engineering |
| 143 | Forensic engineering | Engineering |
| 144 | Remote sensing | Engineering |
| 145 | Marine engineering | Engineering |
| 146 | Simulation | Engineering |
| 147 | Mining engineering | Engineering |
| 148 | Nuclear engineering | Engineering |
| 149 | Biomedical engineering | Engineering |
| 150 | Atmospheric sciences | Environmental science |
| 151 | Meteorology | Environmental science |
| 152 | Climatology | Environmental science |
| 153 | Environmental resource management | Environmental science |
| 154 | Environmental planning | Environmental science |
| 155 | Agricultural science | Environmental science |
| 156 | Waste management | Environmental science |
| 157 | Agronomy | Environmental science |
| 158 | Horticulture | Environmental science |
| 159 | Hydrology | Environmental science |
| 160 | Soil science | Environmental science |
| 161 | Environmental protection | Environmental science |
| 162 | Ecology | Environmental science |
| 163 | Agroforestry | Environmental science |
| 164 | Water resource management | Environmental science |
| 165 | Geomorphology | Environmental science |
| 166 | Forestry | Environmental science |
| 167 | Earth science | Environmental science |
| 168 | Oceanography | Environmental science |
| 169 | Fishery | Environmental science |

| ID | $2^{nd}$ level | $1^{st}$ level |
|---|---|---|
| 170 | Environmental health | Environmental science |
| 171 | Regional science | Geography |
| 172 | Economic geography | Geography |
| 173 | Geodesy | Geography |
| 174 | Physical geography | Geography |
| 175 | Cartography | Geography |
| 176 | Petrology | Geology |
| 177 | Mineralogy | Geology |
| 178 | Paleontology | Geology |
| 179 | Crystallography | Geology |
| 180 | Archaeology | History |
| 181 | Ancient history | History |
| 182 | Genealogy | History |
| 183 | Metallurgy | Materials science |
| 184 | Composite material | Materials science |
| 185 | Ceramic materials | Materials science |
| 186 | Nanotechnology | Materials science |
| 187 | Polymer science | Materials science |
| 188 | Combinatorics | Mathematics |
| 189 | Discrete mathematics | Mathematics |
| 190 | Pure mathematics | Mathematics |
| 191 | Algebra | Mathematics |
| 192 | Statistics | Mathematics |
| 193 | Mathematics education | Mathematics |
| 194 | Actuarial science | Mathematics |
| 195 | Mathematical analysis | Mathematics |
| 196 | Applied mathematics | Mathematics |
| 197 | Topology | Mathematics |
| 198 | Calculus | Mathematics |
| 199 | Mathematical optimization | Mathematics |
| 200 | Arithmetic | Mathematics |
| 201 | Geometry | Mathematics |
| 202 | Psychiatry | Medicine |
| 203 | Orthodontics | Medicine |
| 204 | Dentistry | Medicine |
| 205 | Medical emergency | Medicine |
| 206 | Emergency medicine | Medicine |
| 207 | Ophthalmology | Medicine |
| 208 | Optometry | Medicine |
| 209 | Endocrinology | Medicine |
| 210 | Internal medicine | Medicine |
| 211 | Nursing | Medicine |
| 212 | Family medicine | Medicine |
| 213 | Intensive care medicine | Medicine |
| 214 | Radiology | Medicine |

| ID | $2^{nd}$ level | $1^{st}$ level |
|---|---|---|
| 215 | Nuclear medicine | Medicine |
| 216 | Physical therapy | Medicine |
| 217 | Physical medicine and rehabilitation | Medicine |
| 218 | Cancer research | Medicine |
| 219 | Oncology | Medicine |
| 220 | Medical education | Medicine |
| 221 | Gerontology | Medicine |
| 222 | Virology | Medicine |
| 223 | Immunology | Medicine |
| 224 | Pediatrics | Medicine |
| 225 | Veterinary medicine | Medicine |
| 226 | Pathology | Medicine |
| 227 | General surgery | Medicine |
| 228 | Surgery | Medicine |
| 229 | Nuclear magnetic resonance | Medicine |
| 230 | Genetics | Medicine |
| 231 | Cardiology | Medicine |
| 232 | Anesthesia | Medicine |
| 233 | Obstetrics | Medicine |
| 234 | Gynecology | Medicine |
| 235 | Neuroscience | Medicine |
| 236 | Gastroenterology | Medicine |
| 237 | Traditional medicine | Medicine |
| 238 | Physiology | Medicine |
| 239 | Audiology | Medicine |
| 240 | Urology | Medicine |
| 241 | Andrology | Medicine |
| 242 | Dermatology | Medicine |
| 243 | Anatomy | Medicine |
| 244 | Theology | Philosophy |
| 245 | Aesthetics | Philosophy |
| 246 | Engineering ethics | Philosophy |
| 247 | Epistemology | Philosophy |
| 248 | Environmental ethics | Philosophy |
| 249 | Astronomy | Physics |
| 250 | Astrophysics | Physics |
| 251 | Molecular physics | Physics |
| 252 | Chemical physics | Physics |
| 253 | Quantum electrodynamics | Physics |
| 254 | Quantum mechanics | Physics |
| 255 | Seismology | Physics |
| 256 | Geophysics | Physics |
| 257 | Particle physics | Physics |
| 258 | Nuclear physics | Physics |
| 259 | Atomic physics | Physics |

| ID | $2^{nd}$ level | $1^{st}$ level |
|---|---|---|
| 260 | Classical mechanics | Physics |
| 261 | Mathematical physics | Physics |
| 262 | Theoretical physics | Physics |
| 263 | Condensed matter physics | Physics |
| 264 | Optics | Physics |
| 265 | Biophysics | Physics |
| 266 | Computational physics | Physics |
| 267 | Statistical physics | Physics |
| 268 | Thermodynamics | Physics |
| 269 | Medical physics | Physics |
| 270 | Engineering physics | Physics |
| 271 | Acoustics | Physics |
| 272 | Law | Political science |
| 273 | Public administration | Political science |
| 274 | Clinical psychology | Psychology |
| 275 | Psychotherapist | Psychology |
| 276 | Social psychology | Psychology |
| 277 | Developmental psychology | Psychology |
| 278 | Pedagogy | Psychology |
| 279 | Cognitive psychology | Psychology |
| 280 | Applied psychology | Psychology |
| 281 | Psychoanalysis | Psychology |
| 282 | Criminology | Psychology |
| 283 | Cognitive science | Psychology |
| 284 | Religious studies | Sociology |
| 285 | Social science | Sociology |
| 286 | Gender studies | Sociology |
| 287 | Socioeconomics | Sociology |
| 288 | Media studies | Sociology |
| 289 | Ethnology | Sociology |
| 290 | Anthropology | Sociology |
| 291 | Demography | Sociology |

Table 1.17: Fields of Study of Conversion Table

Note: Conversion table between the second level (292 sub-disciplines) and the first level (19 disciplines) of fields of study. The table also reports node IDs used in the knowledge space (Figure 1.3).

# Chapter 2

# Collaboration and Interdisciplinarity: the Role of Tight Networks in Research Diversification

***Joint with Magda Fontana, Martina Iori and Valerio Leone Sciabolazza***

## 2.1 Introduction

As teamwork rises in science, so does the importance of collaboration networks on researchers' behavior and outcomes (Wuchty et al., 2007). It has been argued that the shift to larger research teams allows scientists to tackle complex problems that require interdisciplinary solutions (Leahey, 2016; Rhoten and Parker, 2004). However, the effects of increasing scientific collaboration is neither straightforward nor homogeneous and it depends on the structure of the established social network. On one hand, researchers may take part in close collaborations between a small group of colleagues, which has been linked with better performance in risky contexts and more disruptive works (Lindenlaub and Prummer, 2021; Wu et al., 2019). On the other hand, scholars may favor looser connections between a larger number of colleagues, which has been shown to provide advantages to developing already established ideas in more stable environments (Lindenlaub and Prummer, 2021; Wu et al., 2019). This evidence suggests that a researcher collaboration structure may act either as a catalyst for disparate and new ideas or as a driver of specialization and productivity. It follows that other things being equal, different kinds of network structures may be associated to different levels of diversity of research subjects and activities an individual engages with. Therefore, the costs and benefits given by different collaboration structures are of particular relevance to the knowledge production process and the career paths scholars take.

In this paper, we study the interplay between different structures of research collaboration and scholars' research portfolio diversification. We investigate how the closeness of researchers' collaborators is related to the level of specialization of their publications. Our focus is on identifying which kind of team structure facilitates the development of interdisciplinary projects. We expect that researchers adjust the diversification of their output in response to the costs and benefits linked to how tightly connected is a research group. Tight connections may influence specialization through two mechanisms: it may increase peer pressure and the need for conformity among researchers; or it allows for better information flow which in turn allows investigators to tap in a larger pool of knowledge. This question is related to the literature that investigates how social networks affect academic outcomes and scientific process (Ductor et al., 2014).

This paper contributes to the empirical study of social interactions in academia. Specifically, we provide evidence of the importance of social networks in the academic workplace. Recently, studies document the role of social ties in the academic job search, promotion and individual performance (Lindenlaub and Prummer, 2021), but the relationship between the collaboration network structure and individual researchers' specialization choices is still not well understood. Our study assesses the role of co-authorship ties on specialization and provide suggestive evidence of risk aversion as a mechanism through which network effects operate in an academic setting.

Furthermore, we provide evidence for the role of social interactions in the adoption of interdisciplinarity by academics. There is a growing body of empirical work that analyses the factors that hinder the pursuit of a more interdisciplinary research agenda. These works mainly focus on the funding constraints (Sun et al., 2021), institutional barriers (Rafols et al., 2012) and the quality and impact of interdisciplinary research (Leahey et al., 2017). This literature highlights the importance of teamwork for interdisciplinarity, yet little attention is given to the costs associated to different social network structures. Our work complements this literature by demonstrating which team structure is associated to more interdisciplinary publications.

In order to examine the role of social ties on researchers' agendas, I use a novel and unique dataset of publications by researchers affiliated with the University of Florida (UF) in the period 2008-2013 to construct their co-authorship networks. The

University of Florida is an interesting case study, given that it is a top-ranked public research university in the United States, employing more than 5,000 researchers in all fields of study and with a research budget of around $1 billion annually. Our empirical strategy relies primarily on the use of individual and year fixed effects. We exploit the variation in their collaborations patterns in each year to examine their relationship with the level of interdisciplinary of individuals' publication records. Specifically, we look at the changes in the number of distinct co-authors of each researcher - i.e. their *degree centrality* - and in the proportion of those colleagues that collaborated among each other - their *clustering coefficient*. Then we connect these network features to the level of disciplinary diversification of researchers' publications in that year. We complement this empirical strategy with an heterogeneity analysis to elucidate the mechanisms that drive our results.

We find that researchers that collaborate in highly clustered groups publish in a less diverse pool of disciplines. We also find evidence that an increase the number of co-authors amplifies the effects of close-knit teams. These results are robust to different measures of interdisciplinarity. Furthermore, we show that the effects of looser networks do not persist from one year to another. These findings support the idea that the tightness of the social network in which researchers embed themselves is systematically related to the interdisciplinarity of their scholarly output. This is in line with theoretical models that predict that tighter collaborations generate more peer pressure and are thus less advantageous in risky and uncertain projects.

To shed further light on the mechanisms at play, we explore heterogeneity in our results due to different individual and co-author attributes. We find that the magnitude of the effects of tight knit groups is less important for man and more important for women. These findings suggest that there is a gender dynamic at play. Furthermore, we show that the effects of tighter networks are stronger for research groups with relatively more newcomers and with at least one woman. Our interpretation is that greater gender diversity and the presence of newcomers increase the inflow of new ideas and methods to teams, which in turn raise the uncertainty of the research outcome. Together, these results suggest that our main results are driven by risk aversion.

In our final set of analysis, we assess whether there are differences in the social network effects by researchers' modal field of study. Although collaborating with a more tight group seems to discourage interdisciplinarity to a certain degree across the board, we find evidence of important heterogeneity across fields: the complementary effect of degree on clustering seems to be relevant mostly on hard science and life science disciplines, while the direct effect of clustering seems to be stronger in social sciences and medicine. These differences suggest that the nature of the research activities and the circumstances in which research is done may play an important role. For example, the costs and benefits of collaboration may be different in more capital intensive fields or in fast-pace knowledge domains, thus creating different teamwork dynamics.

Given that the formation of collaboration teams is the outcome of strategic decisions, it is hard to establish a causal relationship between network structure and researcher specialization. Thus, we conducted a series of additional robustness checks to further corroborate our findings. To assess the robustness of our results to alternative network measures, we analyze the effects of repeated collaborations and construct the co-authorship network across different time windows. Results are fairly robust to alternative measures. We also explore different sources of heterogeneity using alternative discipline and institutional affiliation and observe the same patterns of disciplinary differences. Finally, to emphasize the robustness of our results, we test alternative specifications for the dynamic panel estimations and different modelling strategies for our count dependent variable. The main results hold across all specifications.

The rest of the paper proceeds as follows. Section 2 discusses the related literature and conceptual framework. Section 3 presents data, measures and empirical strategy. Section 4 discusses the results. Section 5 concludes.

## 2.2 Conceptual Framework

*Related Literature.* Our work contributes to the strands of literature that analyze the role of network structures in different social-economic outcome. We see the social interaction between researchers captured by co-authorship networks as part of the scholars' social capital, i.e. a knowledge-based resource that increases the performance of those who collaborate (Li et al. 2013).

Recently papers propose theoretical models to explain the dynamics within the emergence of collaboration networks in academia and their outcomes. Anderson and Richards-Shubik, 2021 estimate a structural network model where the researchers strategically choose to collaborate on projects based on, among other variables, the benefits and costs spillovers coming from networks: the negative spillover coming from working in multiple projects and a positive spillover coming for information increased flows between projects. Furthermore, researchers face the usual individual costs of communication and coordination. They find that coauthors increase output, network spillovers have only modest effects and that generalists do not decrease coordination costs.

In the same vein, Lindenlaub and Prummer, 2021 propose a theoretical model connecting researcher's network structure to productivity. In this model, researchers choose their effort level based on how loose or tight are their network structures. In their setup, groups with more coauthors have increased information flows, which allows researchers to increase effort in projects that are more valuable; while a highly clustered group increases peer pressure, which in turn increases effort. They find that more loosely connected groups with many coauthors have better performance in highly uncertain environment and that the opposite is true for highly clustered groups with less coauthors. They also find that number connections and tightness of networks are complementary factors: more coauthors increase the peer pressure effect.

We follow these efforts by empirically exploring the predictions that come from Anderson and Richards-Shubik, 2021 and Lindenlaub and Prummer, 2021 model adapted to the context of specialization and interdisciplinarity. We investigate if having a higher degree increases the flow of information between collaborators, which would increase the opportunities to recombine knowledge from different disciplines. We also test how groups characterized by high clustering are associated with their level of interdisciplinarity. researchers to work in a lower. Lastly, we are also interested in testing the prediction that higher degree and clustering are complementary in their association with author's interdisciplinarity.

Different from the efforts that focus on the paper or patent level, we take an individual researcher perspective. Instead of assuming individual's specialization as fixed, we allow it to vary over time and study the evolution of a scholar's interdisciplinarity on a yearly basis. Looking at the interdisciplinarity of researchers' production over time enables us better understand the association between the decision to undertake more diverse projects and the collaboration network structure.

*Specialization measures.* We first introduce our interdisciplinarity measures. Instead of taking individual specialization as given by it's departmental affiliation, we assume that individuals' specialization evolves over time. We look at three different dimensions of specialization. It is natural to start by counting the field of study of papers published by researchers in each year. We define our first measure as the number of different subdisciplines a researcher $i$ published in year $t$:

$$V_{it} \equiv \sum_{s \in F} 1, \tag{2.1}$$

where $F$ is the set of all fields of study assigned to author's publications in year $t$

and $s$ denotes a unique field in this set. This measure follows more closely variations in multidisciplinarity rather than interdisciplinarity.

However, the number of fields in which one publishes does not capture all aspects disciplinary diversity. In particular, it does not capture differences in the dispersion of publications in each field. In order to measure the evenness of distribution of publications in different disciplines, we look at the share of publications in each fields of study in a year and calculate the Hirschman-Herfindahl (HH) specialization index. In order to make it comparable with our other measures of interdisciplinarity, we take the unit complement (1 - HH) to define our dispersion index. It is given by:

$$G_{it} \equiv 1 - \sum_{s \in F} (N_s / |F|)^2, \tag{2.2}$$

where $N_s$ is the frequency of field of study $s$ assigned to researcher's publications in year $t$ and $|F|$ is the total number of fields of study assigned to papers published by the same researcher in the same year. This dispersion index is is a real number bewtween 0 and 1 and captures how focused in a discipline an author was in each year, e.g. it assumes value 0 if all researcher' publication that year are in a single field.

We also quantify interdisciplinarity by taking into account the similarity between fields. We do it by using a measure of proximity between disciplines in the idea space. This proximity measure was collected from MAG and is based on the co-occurrence of fields of study in a same paper. We then define our measure of interdisciplinarity based on dissimilarity between fields as the average distance between disciplines in which authors have published, normalized by the total number of unique disciplines in a given year. Formally, we define it as

$$D_{it} \equiv \frac{1}{V_{it}(V_{it} - 1)} \sum_{\substack{r,s \in F \\ r \neq s}} (1 - p_{rs}), \tag{2.3}$$

where $V_{it}$ is the number o distinct disciplines as defined in equation X and $p_{rs}$ is the proximity between fields $r$ and $s$. As proximity is a real number between 0 and 1, our dissimilarity index is also define between this bounds. Note that this measure is not defined for years in which researchers published in only one field and thus are treated as missing values when researcher production is monodisciplinary.

Since the previous measure do not take into account how the fields of study are distributed, we calculate the Shannon entropy - a measure that captures the evenness of the distribution of the disciplines published by author $i$ in year $t$ and is traditionally used as a measure of diversity in information theory. The Entropy measure is given by:

$$E_{it} \equiv \sum_{s \in F} f_s \log f_s, \tag{2.4}$$

where $f_s$ in the relative frequency of field $s$ in publications by the researcher of interest at year $t$.

The Integration Score, also known as diversity index is defined as combination of Variety, Balance and Disparity. This measure synthesizes three different dimensions of interdisciplinarity, taking into account fields' frequency, distribution and similarity. It is given by:

$$R_i \equiv \sum_{\substack{r,s \in F \\ r \neq s}} (1 - p_{rs}) f_r f_s, \tag{2.5}$$

Novelty Index is a measure of atypical combination of fields. Uses the distribution of the proximities between fields where researcher published, takes 10th percentile of that distribution (left tail tendency of distance between disciplines). It is given by:

$$U_i \equiv 1 - 10(p), \tag{2.6}$$

*Network measures.* We construct a co-authorship network where two authors have a link in the network $G$ if they have at least one joint publication in the year $t$. Formally, we define $g_{ij,t} \in \{0, 1\}$ as the status of academic collaboration between authors $i$ and $j$. When $g_{ij,t} = 1$, the two authors published at least one paper together that year and $g_{ij,t} = 0$ when there was no collaboration between the pair in that period. We define $G$ as the network yielded by the collection of authors the respective co-authorship links.

We define the degree of researcher $i$ as the number of distinct coauthors of a researcher $i$ over year $t$. Formally, it is given by:

$$d_{it} = \mid j : g_{ij,t} = 1 \mid . \tag{2.7}$$

When a researcher does not have any publication in year $t$, degree is treated as missing.

Our main network variable is the local clustering coefficient. The clustering coefficient of author $i$ is a measure of the overlap between links of different co-authors, i.e. how many co-authors of a researcher are co-authors themselves. It captures the tightness of connection amongst researcher $i$'s team members in year $t$. It is given by

$$CC_{it} = \frac{\sum_{j \neq i; k \neq j; k \neq i} g_{ij,t} g_{ik,t} g_{jk,t}}{\sum_{j \neq i; k \neq j; k \neq i} g_{ij,t} g_{ik,t}}. \tag{2.8}$$

Clustering coefficient is undefined for researchers with less than two co-authors.

Is clustering the only appropriate measure of network tightness? Are results robust to a different measure of tightness? To test this we introduce strength (repeated collaborations) as another measure of tight network structure.

Strength: normalized average strength across all ties over period t. Strength of ties measure number of papers written with a co-author. Number papers co-authored: $n_{ij,t}$. Then:

$$s_{it} = \frac{1}{d_{it}} \sum_{j: g_{ij,t}=1} n_{ij,t} \tag{2.9}$$

Normalizing $s_{it}$ by dividing by number of publications, in order to capture time spent between co-authors (Ductor, 2020).

$$\bar{s}_{it} = \frac{s_{it}}{P_{it}} \tag{2.10}$$

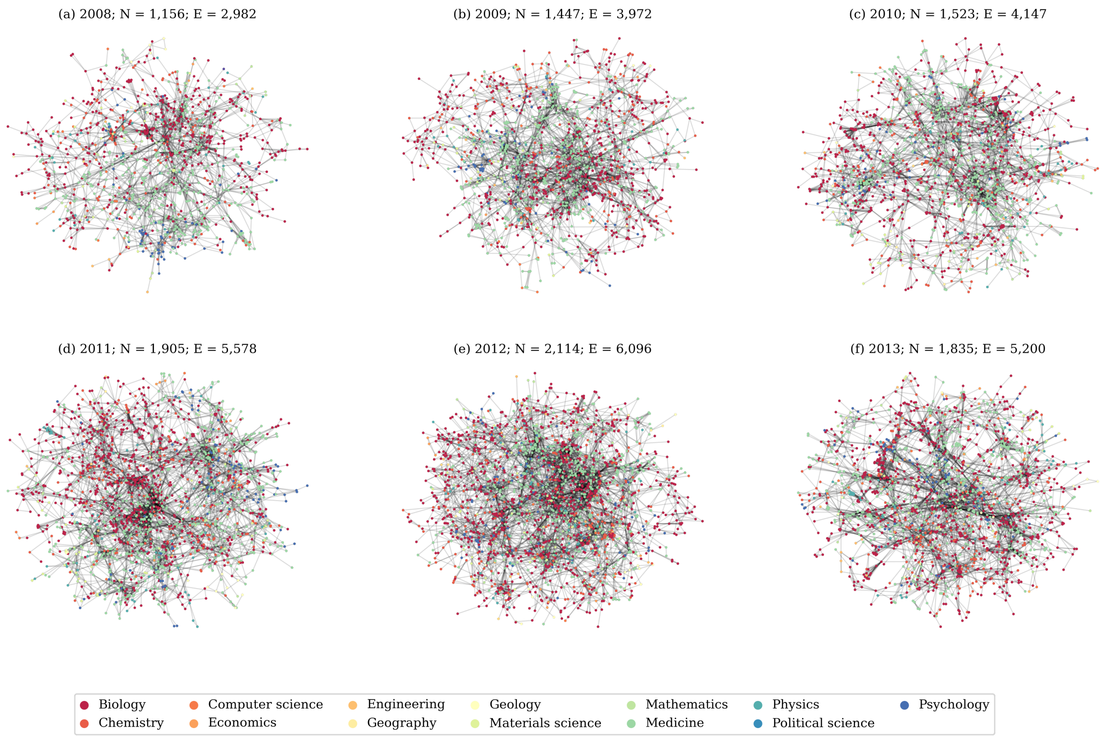
## 2.3 Data

Our sample is drawn from a novel and unique database that tracks individual characteristics and academic publications of researchers affiliated to the University of Florida (UF) from 2008 to 2013. We connect UF's registry office data with bibliometric information from Crossref, Lens and Elsevier Scopus databases, creating a publication record for researchers active in this period[1]. Our original database contains information on 23,926 articles and 6,105 researchers. Using this information we map the collaborations of scholars within UF and construct the university co-authorship networks in each year and construct a panel containing our node (individual researcher) network measures. Figure 2.1 represents the main component of yearly collaboration among UF researchers.

Then, we supplement our paper-level data with field of study classification retrieved from Microsoft Academic Graph (MAG) database The classification scheme implemented by MAG is a hierarchical classification that identifies 19 disciplines

[1] More details regarding the data collection and disambiguation efforts are provided in Appendix

Figure 2.1: Coauthorship Networks from University of Florida researchers by Year




The graphs (a) to (f) depict the giant component of the yearly cross-sectional coauthorship networks formed by University of Florida's researchers from 2008 to 2013. Each node represents a researcher and two researchers are connected if they shared the authorship of a paper in that year. N represents the number of nodes and E the number of edges in the respective networks. The color of the nodes indicates the researchers' modal field of study.

(first level) and 292 sub-disciplines (second level) at the first two levels of classification. Each paper may be tagged with more than one discipline/field based on the full-text of the papers[2]. We use the fields of study information collected from MAG to calculate the time-varying level of interdisciplinarity of the researchers in our sample. Figure 2.1 nodes are colored according to the researchers' modal fields analyzed at the first level.

In order to be able to analyze the evolution of collaboration patterns, we restrict our attention to researchers who had at least two coauthors in at least two years in the period 2008-2013. The former condition is necessary to be able to calculate our main network measure while the latter is due to our empirical strategy that is based on individual fixed effects and thus require within-individual variation. Both requirements will be discussed at length in the following sections. Furthermore, to be able to make meaningful comparisons between different interdisciplinarity measures, we focus only in years where researchers published in more than one subfield. The

[2]For more details on MAG classification scheme, see (cite Sinha, 2019)

rationale for doing so has to do with the different interdisciplinary measures that we use in our analysis: some of them are not defined when researcher's academic production is monodisciplinary. The focus on investigators with these characteristics requires us to exclude a large proportion of the researchers in our original database - our final sample includes 2,446 individuals (40% of the original 6,105 researchers).

Moreover, we analyze only collaborations inside University of Florida. The focus on co-authorship within one institution can be justified both on theoretical and pragmatical ground. First, we believe that close collaborations on the job in a same institution are the ones that are more relevant to the mechanisms we are investigating: scholars from the same department arguably spend more time together and have more meaningful social interactions. From a practical standpoint, we do not have access to disambiguated data from coauthors outside UF, which means that information regarding their publication record and personal characteristics are not acessible.

Table 2.1 provides summary statistics of the variables included in the analysis. Column 1 provides the mean value of the variables, column 2 the median, column 3 the standard deviation, column 4 the minimum values and column 5 the maximum values. Each variables is grouped in panels. In panel A we report the descriptive statistics for our dependent variables, researchers' specialization measure in a year. In panel B we report our main explanatory variables, the network measures of each individual. Panel C reports statistics at the individual level and Panel D reports statistics regarding research team characteristics.

We draw attention to distinctive characteristics of the data. First, we observe that the mean values of our specialization measures are relatively different among themselves. This is to be expected given that the measures capture different aspects of specialization and interdisciplinarity. Second, we observe that the mean researcher in our sample collaborate with 6 other investigators affiliated to the University of Florida and publish 4 papers a year. This rather high numbers is explained by the fact that our analysis excludes researchers with lower productivity and that do not collaborate often. Third, we note that even though women are only 27% of the sample, 73% of the collaborations in a year had the presence of a woman, which indicates that female researchers in our are particularly productive and well connected.

Figure 2.2 shows the relationship between our two main network measures, degree and clustering coefficient, and our main specialization measure, the number of fields of studies. We observe that there is a a positive and significant correlation of 0.46 between researchers' average number of fields and average degree. Moreover, we

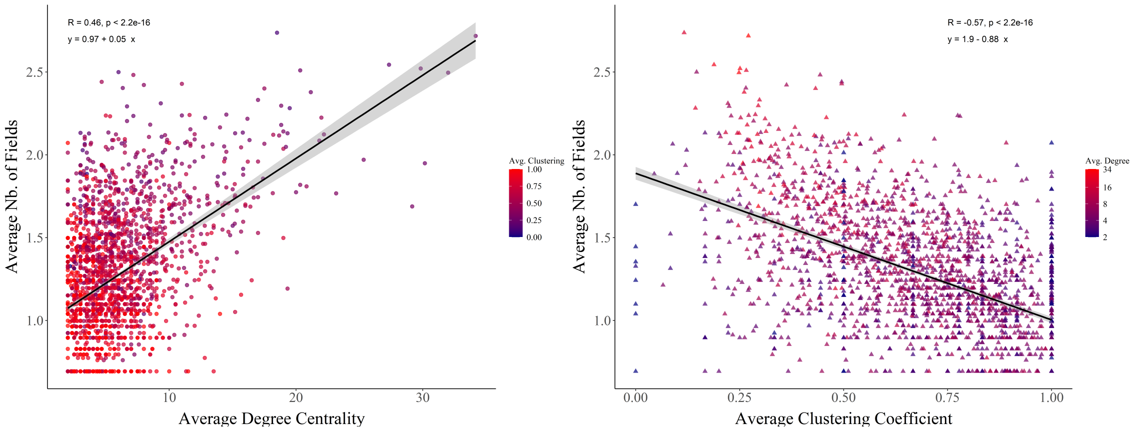
Table 2.1: Descriptive Statistics

| | Mean | Median | SD | Min | Max |
|---|---|---|---|---|---|
| ***Panel A - Specialization Measures*** | | | | | |
| Nb. of Fields | 4.27 | 4 | 2.40 | 2 | 21 |
| Shannon Entropy | 1.24 | 1.20 | 0.46 | 0.38 | 2.82 |
| Dispersion Index | 0.67 | 0.67 | 0.13 | 0.22 | 0.94 |
| Avg. Field Dissimilarity | 0.44 | 0.45 | 0.13 | 0.07 | 0.89 |
| Integration Score | 0.29 | 0.29 | 0.11 | 0.04 | 0.62 |
| Novelty Index | 0.54 | 0.54 | 0.18 | 0.07 | 0.93 |
| ***Panel B - Network Measures*** | | | | | |
| Degree | 6.15 | 5 | 4.88 | 2 | 43 |
| Clustering Coefficient | 0.68 | 0.68 | 0.32 | 0 | 1 |
| Strength | 0.51 | 0.46 | 0.30 | 0.05 | 1 |
| ***Panel C - Individual Characteristics*** | | | | | |
| Gender (Female=1) | 0.27 | 0 | 0.44 | 0 | 1 |
| Nb. of publications | 4.19 | 3 | 3.92 | 1 | 66 |
| H-index | 5.36 | 4 | 5.28 | 0 | 54 |
| Employment type (Faculty=1) | 0.91 | 1 | 0.29 | 0 | 1 |
| Nb. of coauthors | 17.72 | 12 | 19.05 | 0 | 258 |
| Nb. of coauthors within UF | 6.58 | 5 | 5.37 | 2 | 55 |
| Total nb. of coauthors | 61.82 | 42 | 67.69 | 2 | 1,307 |
| Total nb. of coauthors within UF | 17.82 | 14 | 14.59 | 2 | 125 |
| ***Panel D - Group Characteristics*** | | | | | |
| Nb. of new collaborators | 3.59 | 3 | 3.62 | 0 | 39 |
| Coauthored with a woman | 0.73 | 1 | 0.44 | 0 | 1 |
| Total nb. of female coauthors | 5.39 | 4 | 5.63 | 0 | 48 |

*Notes*: The sample includes a panel of 2,446 researchers in University of Florida from year 2008 to 2013. Panel A shows descriptive statistics of selected specialization measures. Panel B shows descriptive statistics for the researcher's network measures: degree centrality, clustering coefficient and normalized average strength. Panel C presents descriptive statistics for measures of productivity, gender, employment type and collaboration patterns assessed at the individual researchers level. Panel D show descriptive statistics of the researcher's group of collaborators in each year. Statistics are calculated at the researcher-year level.

observe a negative and significant correlation of -0.57 between researchers' average number of fields and average clustering coefficient. The scatter plots of figure 2.2 and linear regression lines allow us to confirm visually that the association between our variables of interests are those we expect from our theoretical framework.

Figure 2.2: Scatter plots of Number of Fields of Study on Degree Centrality and Clustering Coefficient




The scatter points represents the average number of fields of study associated to publications, average degree centrality and average clustering coefficient of the researchers in our sample over the period 2008-2013. The lines and gray areas represent the fitted values and 95% confidence intervals of regressions of the average number of fields on the average degree and clustering. The correlation coefficients, p-values and the fitted regression equations are reported on the top of the respective scatter graphs. The color of the dots represents the researchers' corresponding network measures - bluer dots represent lower values while redder dots represent higher values.

## 2.4 Empirical strategy

Our estimation strategy is based on a two-way fixed effects model. The aim is to examine the extent to which a variation of the network structure of researcher $i$ at year $t$ is related to the level of interdisciplinarity of her publications in that period. To this purpose, we run linear regressions on the data using the following equation:

$$IDR_{it} = \beta_1 Degree_{it} + \beta_2 Clustering_{it} + \beta_3 (Degree_{it} \times Clustering_{it}) + \alpha_i + \gamma_t + \varepsilon_{it} \tag{2.11}$$

where $IDR_{it}$ is agent $i$'s interdisciplinarity at year $t$, $Degree_{it}$ and $Clustering_{it}$ are respectively $i$'s degree centrality and $i$'s local clustering coefficient in year $t$, $\alpha_i$ is an individual fixed effect and $\gamma_t$ is a time fixed effects.

By virtue of the variables $\alpha_i$ and $\gamma_t$, this model identifies the effects on the variation of interdisciplinarity, at the individual level, across years. These effects are those determined by the degree centrality and clustering coefficient. Introducing the variable $\alpha_i$ and $\gamma_t$ allow us to control for unobserved time-invariant researchers' characteristics at the individual level, and specific trends at UF in a given year.

If it is true that you take more risks when you have high degree centrality and low clustering, then we should find that $\beta_1 > 0$ and $\beta_2 < 0$. At the same time, since we expect that you take lower risks when you have high degree centrality and high clustering coefficient, we will see that $\beta_3 < 0$.

Then we ask whether clustering is the appropriated network measure to capture tightness of collaborations. In order to evaluate if our results are robust to the introduction alternative network measures, we ran the following regression:

$$\begin{aligned} IDR_{it} = \quad & \beta_1 Degree_{it} + \beta_2 Clustering_{it} + \beta_3 (Degree_{it} \times Clustering_{it}) + \\ & + \beta_4 Strength_{it} + \beta_5 (Degree_{it} \times Strength_{it}) + \\ & + \beta_6 (Degree_{it} \times Strength_{it} \times Clustering_{it}) + \alpha_i + \gamma_t + \varepsilon_{it} \end{aligned} \tag{2.12}$$

where strength is the normalized average strength of the ties between coauthors. We believe that strength is an alternative measure of tightness in scholarly collaboration, thus we expect that you take lower risks when we have repeated interactions with our coauthors $\beta_4 < 0$ and $\beta_5 < 0$. If clustering is not an appropriate measure of network tightness, we would see a qualitative change in $\beta_2$ and $\beta_3$. Furthermore, we ask whether there are any costs or benefits associated to having an extremely tight research group (i.e., having high degree, clustering and strength). We can answer this question through the estimate of coefficient $\beta_6$.

Lastly, we ask whether the network effects are persistent over time. For this scope, we consider an alternative specification based on a dynamic panel model. We add a set of lagged variables to investigate the association of past collaboration patterns with researchers' yearly interdisciplinarity. Then we add a lagged dependent variable and estimate the following model:

$$\begin{aligned} IDR_{it} = \quad & \beta_1 Degree_{it} + \beta_2 Clustering_{it} + \beta_3 (Degree_{it} \times Clustering_{it}) + \\ & + \delta_1 Degree_{it-1} + \delta_2 Clustering_{it-1} + \delta_3 (Degree_{it-1} \times Clustering_{it-1}) + \\ & + \delta_4 IDR_{it-1} + \alpha_i + \gamma_t + \varepsilon_{it} \end{aligned} \tag{2.13}$$

The inclusion of lagged network variables controls for the effect of network structures from the previous year. If there are persistent effects of networks, we would see that the coefficients $\delta$ would be significant.

## 2.5 Results

In this section, we present the empirical results of the article. This section is divided in two parts. In the first part, we present the OLS estimates of equation 2.11, provide an interpretation of the results and present a number of robustness checks for our findings. Finally, in the third part we attempt to explain the mechanisms responsible for our finding by exploring heterogeneity in the results driven by various attributes of individual researchers and their team of collaborators.

### 2.5.1 Main results

Table 2.2 reports the results of regression of researcher's number of fields on their network measures. We regress the network statistics on the number of disciplines in which a researcher publishes in a year. Across all estimates we take both individual and year fixed effects and cluster standard errors around individual researchers. From columns (1) to (4) are estimated with OLS and column (5) is estimated using a System GMM approach.

Consistent with previous research and our conceptual framework, an increase in the degree centrality is positively associated with interdisciplinarity of researcher $i$ in year $t$. Our results reveal that collaborating with one more co-author in a year is associated with 4.4% increase in the number of fields of study assigned to authors' publications (Table 2.2, column (1)). In contrast, an increase in the clustering is associated with a decrease in the interdisciplinarity of authors' publications in a year. In average, we can see in column (1) that having fully connected group of coauthors (i.e. clustering coefficient equal to 1) is associated with 24.3% less fields of studies assigned to publications compared to years in each the researchers have totally disconnected coauthors. Furthermore, the interaction term has significant and negative coefficient, which suggests that degree and clustering are complementary - having more coauthors strengthen the effects of tight networks.

To ensure our results are not driven by the choice of time window to calculate our network measures, in column (2) we report the results of our estimates when we consider the coautorship networks across 2 years (t and t-1). The results are qualitatively unchanged, although the point estimate of the clustering variable is substantially lower (implied elasticity is -7.2%, compared to the -24.3% from the main specification). We note that even though we lost one year of observation by using a larger time window, we have a higher number of observations because there is a larger pool of researchers that are active and published with at least 2 coauthors. We conclude from this estimation that results are robust to considering a larger time

Table 2.2: Co-authorship Networks and Specialization

| | Dependent variable: log(Nb. of Fields) | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Degree | 0.043*** | 0.032*** | 0.043*** | 0.036*** | 0.039*** |
| | (0.002) | (0.002) | (0.003) | (0.003) | (0.002) |
| Clustering | −0.278*** | −0.075** | −0.080* | −0.372*** | −0.456*** |
| | (0.028) | (0.028) | (0.038) | (0.038) | (0.034) |
| Degree × Clustering | −0.033*** | −0.025*** | −0.029*** | −0.017** | −0.025*** |
| | (0.005) | (0.005) | (0.008) | (0.006) | (0.006) |
| Strength | | | −0.631*** | | |
| | | | (0.042) | | |
| Degree × Strength | | | −0.076*** | | |
| | | | (0.014) | | |
| Degree × Clustering × Strength | | | 0.057*** | | |
| | | | (0.014) | | |
| $\text{Degree}_{\text{t-1}}$ | | | | −0.003 | −0.004 |
| | | | | (0.002) | (0.003) |
| $\text{Clustering}_{\text{t-1}}$ | | | | −0.053† | −0.004 |
| | | | | (0.028) | (0.036) |
| $\text{Degree}_{\text{t-1}}$ × $\text{Clustering}_{\text{t-1}}$ | | | | 0.009 | −0.004 |
| | | | | (0.006) | (0.005) |
| log(Nb. of $\text{Fields}_{\text{t-1}}$) | | | | | 0.116*** |
| | | | | | (0.029) |
| Researcher Fixed Effects | YES | YES | YES | YES | YES |
| Year Fixed Effects | YES | YES | YES | YES | YES |
| Number of Researchers | 2,446 | 2,899 | 2,446 | 2,180 | 1,951 |
| Observations | 8,245 | 9,729 | 8,245 | 5,452 | 4,974 |
| $R^2$ | 0.675 | 0.739 | 0.714 | 0.731 | 0.735 |
| Adjusted $R^2$ | 0.537 | 0.628 | 0.593 | 0.550 | 0.557 |
| No. of instruments | — | — | — | — | 25 |
| AR1 (p-value) | — | — | — | — | 0.000 |
| AR2 (p-value) | — | — | — | — | 0.257 |
| Hansen-J (p-value) | — | — | — | — | 0.000 |

*Notes*: This table reports estimates of regressions of our models describe in section 3.4. The dependent variable is the logarithm of the number of fields of study in which a researcher published a paper in the year of observation. Columns (1) reports our baseline results of the estimates of equation 2.11 using OLS. Column (2) reports OLS estimates of the same equation but using measures computed with networks across 2 years. Column (3) reports OLS estimates of equation 2.12 by introducing the normalized average strength and its interactions with the other network measures as controls. Column (4) reports OLS estimates of the main equation when including lagged explanatory variables. Column (5) presents estimates using System GMM of a dynamical panel as described in equation 2.13. Exponentiating the coefficients and differencing from one yields numbers interpretable as elasticities. The number of observations varies across columns because network measures are only defined for those with more than 2 coauthors in the observed period and because the lagged and dynamical panel specifications drop observations corresponding to scientists for without publications over the previous periods. All the specifications include researcher and year fixed effects. The p-values of the usual dynamic panel Arellano–Bond test for first-order (AR1) and second-order (AR2) serial correlation and the Sargan–Hansen over-identifying restrictions test are reported in the last column. Standard errors are clustered at the researcher level. Significance levels: † p<0.1; * p<0.05; ** p<0.01; *** p<0.001.

interval to construct the coauthorship networks.

Column (3) displays the effects of controlling for an additional measure of tightness in collaboration: the researcher's normalized average tie strength. We observe that the inclusion of a measure that capture recurring collaborations does not qualitatively change our main results. Once more the clustering coefficient estimate shows that its effect have smaller magnitude after controlling for strength, although being still negative and significant. Furthermore, the estimates of strength and its interaction with degree imply that repeated interactions is associated with a large decrease in the interdisciplinary (implied elasticity is -46,8%) and the higher the number of collaborators, stronger is this effect. Together with the results for clustering, these estimates further indicates increasing the tightness of collaboration is negatively and strongly associated with the level of interdisciplinarity of one's scholarly publications. Perhaps surprisingly, we can see from the positive and significant estimates of the triple interaction between degree, clustering and strength that the negative effects are partially mitigated when researchers collaborate repeatedly with many colleagues that coauthor among each other. Our interpretation is that being part of an extremely tight knit-groups may offset some of the uncertainties related to the consequences to the group of the possible failure of the project.

A possible concern about these results is that we are not controlling for past collaboration patterns. Thus we ask whether the network effects are persistent in time. To test this hypothesis modeled in equation 2.13, we add lagged network variables to our main specification to capture the effects of scholars' collaborations in the previous year (t-1). In column (4) of we show the estimates for the specification with lagged variables. Our main results remain qualitatively unchanged and the lagged variables coefficients are not statistically significant at the conventional 5% level - only lagged clustering variable is negative and significant at 10% level. However, as we can see in the estimations in column (5), when we add the lagged dependent variable to the specification, the point estimates for lagged variables are not significant in any level. Moreover, the coefficient of the lagged dependent is positive and significant. Therefore, there is evidence that researchers are increasing their level of interdisciplinarity over time, but the results do not suggest that network structure effects persist from one year to another.

Overall, the results are in line with our predictions. Working in loosely connected team (high degree and lower clustering) is related to a higher interdisciplinary in researchers' output. In contrast, when researchers publish with a tightly connected team (high degree and high clustering), their output displays lower interdisciplinarity. Our findings suggests that there are substantial obstacles to undertake interdisciplinary academic production when working with tight knit-groups, but at the same time confirms the importance of increasing number of co-authors to enlarge the pool of disciplines available to knowledge recombination. Furthermore, we have suggestive evidence that these effects are mitigate by extremely tight groups and that the collaboration structure effects are not persistent over time.

Table 2.3: Robustness to Alternative Specialization Measures

| | *Dependent variable:* | | | | |
|---|---|---|---|---|---|
| | Shannon Entropy | Dispersion Index | Avg. Field Dissimilarity | Integration Score | Novelty Index |
| | (1) | (2) | (3) | (4) | (5) |
| Degree | 0.034*** | 0.007*** | 0.004*** | 0.005*** | 0.005*** |
| | (0.002) | (0.001) | (0.001) | (0.001) | (0.001) |
| Clustering | −0.248*** | −0.080*** | −0.048*** | −0.060*** | −0.104*** |
| | (0.027) | (0.008) | (0.008) | (0.007) | (0.011) |
| Degree × Clustering | −0.027*** | −0.005*** | −0.004** | −0.004*** | −0.004* |
| | (0.005) | (0.001) | (0.001) | (0.001) | (0.002) |
| Researcher Fixed Effects | YES | YES | YES | YES | YES |
| Year Fixed Effects | YES | YES | YES | YES | YES |
| Number of Researchers | 2,446 | 2,446 | 2,446 | 2,446 | 2,446 |
| Observations | 8,245 | 8,245 | 8,245 | 8,245 | 8,245 |
| $R^2$ | 0.638 | 0.582 | 0.575 | 0.604 | 0.575 |
| Adjusted $R^2$ | 0.485 | 0.405 | 0.395 | 0.437 | 0.394 |

*Notes*: The table reports OLS estimates of model 2.11 using alternative specialization measures. All the specifications include researcher and year fixed effects. Standard errors are clustered at the researcher level. Significance levels: † p<0.1; * p<0.05; ** p<0.01; *** p<0.001.

*Robustness.* A possible concern about these results is that the number of fields in which a researcher published each year is not a good proxy for researcher's interdisciplinarity. To address this concern, we estimate our main equation 2.11 using alternative interdisciplinarity measures as dependent variables. We show the results of these estimation in Table 2.3 and see the same pattern emerges different output variables. In column (1) we estimate the relationship between tight collaborations and the diversity of fields of study measured by the Shannon Entropy. The results are the same: degree is positively associated with the diversity of fields in researcher's publication and clustering is negatively associated with it. In column (2) we measure if the the structure of one's network is associated with publications that are more dispersed across fields. Results remain unchanged. We then turn to investigate in column (3) tightness of one's connection is associated with the distance between disciplines in which they publish. Once again, the results are qualitatively similar to those using only the count of disciplines. In column (4) we use as dependent variable the Integration Score, a composite interdisciplinary index, and in column (5) we use the Novelty index. In both cases, results are the same. We conclude from this analysis that our results are robust to a variety of interdisciplinary measures, which indicates that the association between research group tightness and interdisciplinarity is systematic and consistent with our conceptual framework.

Table 2.4: Robustness to Alternative Model Specifications

| | *Dependent variable:* | | |
|---|---|---|---|
| | log(Nb. of Fields) | Nb. of Fields | Nb. of Fields |
| | (1) | (2) | (3) |
| | *OLS* | *Poisson* | *Neg. Bin.* |
| Degree | 0.043*** | 0.021*** | 0.021*** |
| | (0.002) | (0.002) | (0.002) |
| Clustering | -0.278*** | -0.284*** | -0.284*** |
| | (0.028) | (0.021) | (0.021) |
| Degree × Clustering | -0.033*** | -0.009** | -0.009** |
| | (0.005) | (0.004) | (0.004) |
| Researcher Fixed Effects | YES | YES | YES |
| Year Fixed Effects | YES | YES | YES |
| Observations | 8,245 | 8,245 | 8,245 |
| $R^2$ | 0.675 | 0.666 | 0.666 |
| Pseudo $R^2$ | 0.770 | 0.053 | 0.053 |
| BIC | 24,892.8 | 40,655.4 | 40,656.4 |
| Over-dispersion | — | — | 10,000 |

*Notes*: The table reports estimates of equation 2.11 using alternative econometric models. Estimates in column (1) replicate our main results with dependent variable being the logarithm of the number of fields of study in which a researcher published a paper in the year of observation. Column (2) shows the results stemming from a quasi-maximum likelihood fixed effects Poisson specification with dependent variable being the count of fields of study. Column (3) shows the results from estimating our main specification using a negative binomial fixed effect model. All the specifications include researcher and year fixed effects. Standard errors are clustered at the researcher level. Significance levels: † p<0.1; * p<0.05; ** p<0.01; *** p<0.001.

Our results are also robust to different functional form of our model, as shown in Table 2.4. In column (1) we replicate our main specification based on fixed-effects OLS, where we use the logarithm of the number of fields as a dependent variable. As the number of fields is a count variable, we present also estimates using a fixed-effects Poisson model estimated through quasi-maximum likelihood (QML) approach. Once more, as we can see in column (2), the results are notably close and qualitatively the same. We then proceed to test a negative binomial specification using QML, but the over-dispersion test indicates that mean and variance in our data are the same, favoring the Poisson model specification. These exercises indicate that our results are robust to altenative model specifications.

Furthermore, we assess potential concerns that our results may be biased because we only observe collaborations within University of Florida. We run our preferred specification restricting the sample to only researchers that had higher share of co-authors from inside UF than outside. Estimations from table 2.8 in appendix show that the there are qualitative differences between researchers who coauthored mainly with UF researchers and those who had the majority of collaborations with authors not affiliated to UF. This results holds if we measure collaborations both yearly and in total collaborations.

Taken together, these results suggest that collaborations with loosely connected research team is associated with having a more interdisciplinarity output while collaborating with a tightly connected group is negatively associated with publications with a lower levels of interdisciplinarity. The main results hold across different model specifications, several alternative measures of interdisciplinarity, controlling for alternative network measures and restricting the sample to those researchers with most of their collaborations within UF. To investigate the drivers of our results, we proceed with an heterogeneity analysis inspired by the theoretical framework presented in section 2.2.

### 2.5.2 Heterogeneity analysis

Having established the negative relationship between clustering and interdisciplinarity at the individual researcher level, we turn to examining whether the effects of network structure may vary according to the characteristics of the researcher and their co-authors. To this purpose, we split our database in subsamples defined by individual and research team characteristics, estimate the main equation and compare the estimates across groups.

*Individual characteristics.* In Table 2.5 we report the results for the regression of the main equation 2.11 restricted to the subsamples given by individual characteristics. First, we assess whether there are gender differences that drive our results using as dependent variable the number of fields. In columns (1a) and (1b) we divide our samples by gender and find that qualitatively the results are the same from those in the main results: degree has a positive coefficient and clustering and the interaction term is negative. We test if these coefficients are statistically different using a generalized Hausman test and find that the effect of tight networks is larger for women. This finding may be explained by the difference in collaboration patterns between women and men. As Lindenlaub and Prummer (2020) and Ductor, Goyal and Prummer (2021) document that woman tend to have lower degree and higher clustering coefficient than men.

Next we ask whether our results are different for highly productive researchers who have many collaborations. It may be the case that more prominent scholars have a higher level of social capital that may be "spent" to have a more interdisciplinary production. We evaluate this possibility by comparing the coefficient from star researchers and non star researchers. We define a "superstar" researchers in the upper 10% percentile of the h-index distribution. We present the results for subsample split by productivity in columns (2a) and (2b) of Table 2.5. We find no evidence of heterogeneity by productivity - the interdisciplinarity of star researchers academic output presents the same negative association with tightly connected teams than non-stars. Although the estimates is less precise, we find that there are no qualitative differences between the effects of clustering between star researchers and non-star researchers. These results indicate that productivity does not explains the association between group tightness and specialization.

Then we examine whether the results vary depending on the employment type of the researcher, i.e. if they are considered faculty or non faculty. It may be the case that non-faculty members, who are not considered scholars in the departments and probably are not the principal investigators, have a weaker link between their network structure and the interdisciplinarity of their output. The estimates in column (3a) and (3b) show that this is not the case, we find no heterogeneity by different employment type.

Our heterogeneity analysis based on individual characteristics shows little evidence that status or productivity are drivers of our results. The only group who present significant difference in the estimates for clustering are woman. These results suggests that the social mechanism behind it seems to be linked to the nature of interdisciplinary projects and the interaction between researchers and not their personal characteristics.

*Team composition.* To further investigate heterogeneous effects in our analysis, we examine the extent to which the characteristics of my collaborators affect the relationship between interdisciplinarity and tight networks.

Do the effects of tighter collaborations vary when we face new collaborators? We answer this question by looking at the differences between groups with number of new collaborations above median and those groups below median. We define new collaborators as coauthors that did not produced a paper with the researcher at any

Table 2.5: Heterogeneity by Individual and Group Characteristics

| | Dependent variable: log(Nb. of Fields) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Gender | | Superstar status | | Employment status | | Team Composition | | | |
| | Men | Women | Star | Non-star | Faculty | Non-faculty | ≥3 New Co-authors | <3 New Co-authors | Female Co-author | No Female Co-author |
| | (1a) | (1b) | (2a) | (2b) | (3a) | (3b) | (4a) | (4b) | (5a) | (5b) |
| Degree | 0.041*** | 0.037*** | 0.030*** | 0.054*** | 0.040*** | 0.053** | 0.031*** | 0.091*** | 0.037*** | 0.091*** |
| | (0.003) | (0.006) | (0.005) | (0.003) | (0.003) | (0.016) | (0.004) | (0.018) | (0.003) | (0.014) |
| Clustering | -0.381*** | -0.527*** | -0.357** | -0.340*** | -0.397*** | -0.447** | -0.588*** | -0.214* | -0.457*** | -0.203* |
| | (0.032) | (0.062) | (0.126) | (0.032) | (0.034) | (0.145) | (0.069) | (0.084) | (0.036) | (0.083) |
| Degree × Clustering | -0.021*** | -0.023* | -0.028† | -0.040*** | -0.024*** | -0.051* | -0.022** | -0.063** | -0.022*** | -0.071*** |
| | (0.006) | (0.010) | (0.015) | (0.006) | (0.006) | (0.023) | (0.008) | (0.024) | (0.006) | (0.021) |
| Researcher Fixed Effects | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Year Fixed Effects | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| No. of researchers | 1,898 | 838 | 368 | 2,772 | 1,748 | 260 | 1,842 | 2,011 | 2,354 | 1,376 |
| Observations | 6,131 | 2,332 | 873 | 7,703 | 5,487 | 575 | 3,471 | 3,352 | 6,231 | 2,345 |
| Test: Degree(a)=Degree(b) | 0.45 (0.50) | | 16.55 (0.00) | | 0.58 (0.45) | | 10.87 (0.00) | | 14.92 (0.00) | |
| Test: Clust.(a)=Clust.(b) | 4.41 (0.04) | | 0.02 (0.90) | | 0.11 (0.74) | | 11.89 (0.00) | | 8.01 (0.00) | |
| Test: Interaction(a)=Interaction(b) | 0.01 (0.91) | | 0.52 (0.47) | | 1.27 (0.26) | | 2.67 (0.10) | | 4.91 (0.03) | |

*Notes*: The table reports OLS estimates of model 2.11 for different subsamples. Columns (1a)-(1b) show results for samples split by gender. Columns (2a)-(2b) show results of samples split by productivity, where star is defined as the top 10% of researchers with the highest h-index. Columns (3a)-(3b) show the results for samples split by employment status, where faculty is defined as researchers who are faculty are academic staff in a department in UF while non-faculty are non-academic staff. Columns (4a)-(4b) present results for samples split by the higher/lower presence of collaborations not observed in the years previous than the observation year (threshold given by the median). Columns (5a)-(5b) present results for samples split by the existence of a collaboration with a woman in the year of observation. All the specifications include researcher and year fixed effects. In addition we report the Hausman test for differences in coefficients between each pair of split samples. Standard errors are clustered at the researcher level. Significance levels: † $p<0.1$; * $p<0.05$; ** $p<0.01$; *** $p<0.001$.

year before the year of observation. In columns (4a) and (4b) of Table 2.5 we present the the estimates for the samples divided by the level of new collaborations. We find that the effect of clustering are significantly higher for groups with many new co-authors. Our interpretation is that there is higher uncertainty associated to new collaboration, which leads to researchers assuming less risks and focus in publishing in a less diverse pool of disciplines.

We then ask whether there is heterogeneity by gender diversity of the research team. A more diverse team may increase the flow of novel ideas and increase coordination costs, which in turn raise the uncertainties associated to the success of the projects. Indeed we show in columns (5a) and (5b) of Table 2.5 that the tightness of a researcher team affects the interdisciplinarity of their output more when that group include a woman. These results reinforce the findings in columns (1a) and (1b), which suggest that there is a gender dynamic at play.

On the whole, these results support our conjecture that the main mechanism is driving our results are uncertainty and risk aversion. Group characteristics seems to be more important than individual characteristics. Furthermore, it seems that gender plays an important role in the effects of network tightness.

*Disciplinary differences.* We now explore whether the effects of network structure may vary according to the modal field of researcher. To this purpose, we estimate equation 2.11 considering splitting our researchers by modal field using Seemingly Unrelated Regressions estimations. Researchers' modal field are defined using the whole sample: we look at the publication record of researchers in the 2008-2013 period and identify the discipline in which the researcher published the most. This discipline is considered our modal field. Than we categorize their fields of study in 4 areas of research: Hard Sciences, Social Sciences, Life Sciences and Medicine. The list of fields and their categorization can be found in the Appendix.

In Table 2.6 we report the results of our estimates of the effects of network structure on the number of fields by each research area. We can see that the results are qualitatively similar to the main estimates for researchers in Hard Sciences - column (1), Life Sciences - column (3) and Medicine - column (4): degree has a significant and positive coefficient and there's a negative and significant negative between clustering and number of fields, which is stronger as the number of coauthors increase. Perhaps unexpectedly, we do not find evidence for the complementarity between degree and clustering for social scientists (column (2)). Our interpretation is that more capital intensive fields of study increase the stakes of collaborations and therefore the risk associated to tighter networks. In appendix Table 2.9 we present an alternative division of researchers in disciplines based in college affiliation as a

Table 2.6: Heterogeneity by Researchers' Main Field of Study

| | Dependent variable: log(Nb. of Fields) | | | |
|---|---|---|---|---|
| | Hard Sciences (1) | Social Sciences (2) | Life Sciences (3) | Medicine (4) |
| Degree | 0.069*** | 0.035** | 0.047*** | 0.035*** |
| | (0.007) | (0.012) | (0.003) | (0.003) |
| Clustering | -0.279*** | -0.538*** | -0.324*** | -0.582*** |
| | (0.057) | (0.105) | (0.035) | (0.042) |
| Degree × Clustering | -0.044** | 0.016 | -0.043*** | -0.015* |
| | (0.014) | (0.024) | (0.006) | (0.006) |
| Researcher Fixed Effects | YES | YES | YES | YES |
| Year Fixed Effects | YES | YES | YES | YES |
| Nb. of Researchers | 403 | 118 | 980 | 797 |
| Observations | 1,738 | 561 | 3,927 | 3,240 |

*Notes*: The table reports OLS estimates of model 2.11 for different subsamples. We divide the sample of researchers by main field of study. Researchers' main field of study is defined by grouping individuals modal subfields in 4 scientific areas following the categorization of the European Research Council. Details on the way the researchers were categorized are provided in the Appendix. In column (1) we report results for the sample of researchers in Hard Sciences; in column (2) we report results for researchers in Social Sciences; in column (3) we report results for researchers in Life Sciences; and in column (4) we report the results for researchers in Medicine. All the specifications include researcher and year fixed effects. Standard errors are clustered at the researcher level. Significance levels: † p<0.1; * p<0.05; ** p<0.01; *** p<0.001.

robustness check. Results are similar, although less precise.

## 2.6 Conclusion

This paper explores the relationship between researchers' collaboration network and the interdisciplinarity of their output. We utilize a novel and unique database containing the publication record of and personal information of investigators affiliated to the University of Florida. We use fixed effect estimations to study at the individual level the patterns of collaboration that are more or less associated with interdisciplinarity. We document that loose networks are associated with higher interdisciplinarity and tighter networks are associated with lower interdisciplinarity. These results are robust to a series of checks. We explore heterogeneity in our results to investigate the mechanisms that drive our results. Our analysis suggest that risk aversion is a main factor in explaining the association between network structure and interdisciplinarity.

Taken together with the well-documented rise in collaboration in academia, this findings imply that researchers face increasing complexity regarding interdisciplinary work. Besides the well known institutional bias against interdisciplinarity, the internal dynamic of teamwork may hinder the adoption of a more diverse research agenda. Our results cast doubts on the hope that increase in the number of collaborators will be translated into an increase of interdisciplinary knowledge production. Thus, if we want to accelerate the engagement of scholars interdisciplinary projects, we should take into consideration their collaboration network structure. Still, there is not straightforward answer in how to devise incentives to achieve this goal. Fomenting collaboration between researchers outside their social groups may positively effect their disposition to take more risks, but at the same time it may increase communication costs and diminish the quality of the output.

## Acknowledgements

This research uses data from the Bureau of Economics and Business Research (BEBR) of the University of Florida. We thank the staff of the BEBR and the director Christopher McCarty for their important contribution to this work.

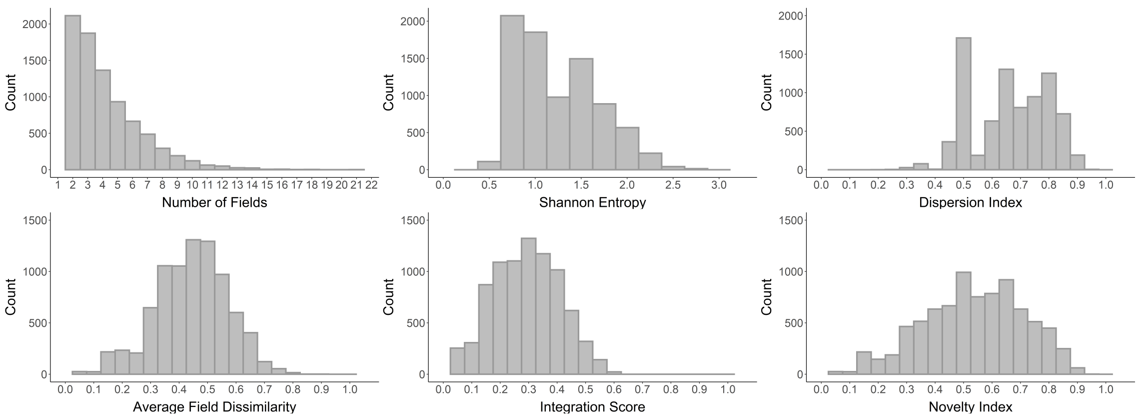
## 2.7 Appendix for Chapter 2

In this appendix we report some additional descriptive statistics and further robustness checks.

### 2.7.1 Additional descriptive statistics

In this section we present additional descriptive statistics of our dataset. In Figure 2.3 we report histograms showing the distribution of each variable measuring interdisciplinarity. In order to have a homogeneous sample and be able to do meaningful comparisons, we exclude the missing values (i.e. observations of years author do not publish or publish monodisciplinary papers). Furthermore, we restrict our sample to only active researcher, defined as those who published in at least two years in the period 2008-2013. Lastly, we exclude observations with missing values in their network measures.

Figure 2.3: Distribution of Specialization Measures

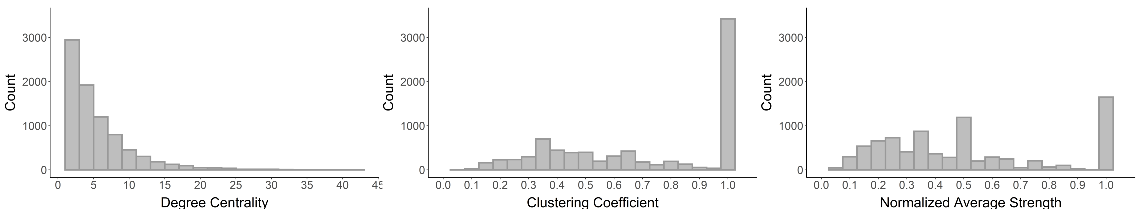



The histograms show the distribution of our measures of specialization. The unit of observation is at the researcher-year level. The sample includes 2,446 researchers and 8,245 observations.

Figure 2.4 we report histograms showing the distribution of each network measure. As can be seen in the distributions of degree centrality and clustering coefficient, we excluded observations of researchers that collaborated with only one other investigator in UF in a year. The reason for that choice is that the clustering coefficient is not defined for researchers who have degree lower than 2.

In Table 2.7 we report the correlation between the main variables used in our regression. Correlations are calculated at the researcher-year level. As we can observe, all correlations have the expected sign and are significant at the 0.1% level.

Figure 2.4: Distribution of Network Measures Measures




The histograms show the distribution of our network measures. The unit of observation is at the researcher-year level. The sample includes 2,446 researchers and 8,245 observations.

Table 2.7: Correlation Matrix

| | Number of Fields | Shannon Entropy | Dispersion Index | Avg. Field Dissimilarity | Integration Score | Novelty Index | Degree | Clustering | Strength |
|---|---|---|---|---|---|---|---|---|---|
| Number of Fields | — | | | | | | | | |
| Shannon Entropy | 0.93*** | — | | | | | | | |
| Dispersion Index | 0.82*** | 0.96*** | — | | | | | | |
| Avg. Dissimilarity | 0.42*** | 0.46*** | 0.43*** | — | | | | | |
| Integration Score | 0.68*** | 0.78*** | 0.78*** | 0.86*** | — | | | | |
| Novelty Index | 0.56*** | 0.65*** | 0.65*** | 0.93*** | 0.93*** | — | | | |
| Degree | 0.50*** | 0.41*** | 0.32*** | 0.17*** | 0.25*** | 0.22*** | — | | |
| Clustering | -0.48*** | -0.45*** | -0.38*** | -0.26*** | -0.33*** | -0.33*** | -0.41*** | — | |
| Strength | -0.58*** | -0.57*** | -0.49*** | -0.32*** | -0.41*** | -0.42*** | -0.44*** | 0.71*** | — |

*Notes*: The table reports the correlation matrix between our specialization and network measures. The unit of observation is the researcher-year. The sample consists of 2446 researchers and 8245 observations. Significance levels: *** p<0.001.

### 2.7.2 Robustness: Restricted Research Samples

As a robustness check we assess potential concerns that our results may be biased because we only observe collaborations within University of Florida. We run our preferred specification restricting the sample to only researchers that had higher share of co-authors from inside UF than outside. Results from table 2.8 show estimates for the sample of researchers that had more than half of collaborations within UF. Given that we do not have access to the publication records of coauthors outside the UF, we cannot extend our analysis including the full collaboration network. However, we can compare the main results restricting the sample to researchers that we observe a higher proportion of their network. Table 2.8 presents the results for the restricted samples. In column (1) we replicate the results of the full sample. In column (2) we restrict our sample to researchers with more than 50% of their collaborations within UF in the year of observation and results do not change. In column (3) we consider an even more strict sample: we exclude researchers that had less than half of their collaborations in the period whole 2008-2013. The results are confirmed even when we consider a sample of researchers that have most of their collaborations within UF.

Table 2.8: Relative Importance of Collaborations within University of Florida

| | Dependent variable: log(Nb. of Fields) | | |
|---|---|---|---|
| | Full Sample | >50% within UF by year | >50% within UF total |
| | (1) | (2) | (3) |
| Degree | 0.043*** | 0.047*** | 0.045*** |
| | (0.002) | (0.005) | (0.007) |
| Clustering | −0.278*** | −0.370*** | −0.395*** |
| | (0.028) | (0.062) | (0.073) |
| Degree × Clustering | −0.033*** | −0.042*** | −0.046*** |
| | (0.005) | (0.008) | (0.010) |
| Researcher Fixed Effects | YES | YES | YES |
| Year Fixed Effects | YES | YES | YES |
| Avg. number of co-authors year | 17.72 | 10.85 | — |
| Avg. number of co-authors total | 61.82 | — | 27.71 |
| Number of Researchers | 2,446 | 1,651 | 552 |
| Observations | 8,245 | 3,184 | 1,605 |
| $R^2$ | 0.675 | 0.755 | 0.632 |
| Adjusted $R^2$ | 0.537 | 0.489 | 0.436 |

*Notes*: The table reports OLS estimates of equation 2.11. The dependent variable is the logarithm of the number of fields of study in which a researcher published a paper in the year of observation. Estimates in column (1) replicates our main results. Estimates in column (2) restricts the sample to researchers with the majority (more than 50%) of their coauthors within University of Florida in the year of observation. Estimates in column (3) restricts the sample to researchers which had the majority of their collaborations within the University of Florida in the whole period 2008-2013. All the specifications include researcher and year fixed effects. Standard errors are clustered at the researcher level. Significance levels: † p<0.1; * p<0.05; ** p<0.01; *** p<0.001.

### 2.7.3 Robustness: Alternative Institutional Affiliation

In this section we further proceed with our analysis by estimating our main equation when considering a sample of researchers affiliated to a specific colleges or academic unit division. This is an alternative division in disciplines to those we use previously that were based on researcher's modal field of publication.

Table 2.9: Heterogeneity by Researchers' College Affiliation

| | Dependent variable: log(Variety) | | | |
|---|---|---|---|---|
| | Engineering College | Humanities & Sciences Colleges | Life Sciences Colleges | Health Sciences Colleges |
| | (1) | (2) | (3) | (4) |
| Degree | 0.069*** | 0.045*** | 0.047*** | 0.034*** |
| | (0.012) | (0.007) | (0.006) | (0.003) |
| Clustering | -0.314** | -0.355*** | -0.285*** | -0.519*** |
| | (0.098) | (0.064) | (0.050) | (0.036) |
| Degree × Clustering | -0.036 | -0.021 | -0.025* | -0.021*** |
| | (0.024) | (0.016) | (0.011) | (0.005) |
| Researcher Fixed Effects | YES | YES | YES | YES |
| Year Fixed Effects | YES | YES | YES | YES |
| Nb. of Researchers | 179 | 209 | 579 | 1,250 |
| Observations | 599 | 945 | 1,878 | 4,187 |

*Notes*: The table reports OLS estimates of model 2.11 for different subsamples. We divide the sample of researchers college affiliation. Researchers' college is defined by grouping individuals' departments in 4 categories of colleges as defined by the structure of University of Florida. Details on the way the researchers were categorized are provided in the Appendix. In column (1) we report results for the sample of researchers in Engineering College; in column (2) we report results for researchers in Humanities and Sciences Colleges; in column (3) we report results for researchers in Life Sciences Colleges; and in column (4) we report the results for researchers in Health Science Colleges. All the specifications include researcher and year fixed effects. Standard errors are clustered at the researcher level. Significance levels: † p<0.1; * p<0.05; ** p<0.01; *** p<0.001.

# Chapter 3

# Economists on the Twittersphere: Analysing Research Diffusion in an Emerging Small World

## 3.1 Introduction

Twitter has become one of the most important means of communication in the world, reaching in the second quarter of 2021 a total of 211 million daily active users (Twitter, 2021). It serves as communication channels between business and costumers, employers and job-seekers, politicians and voters. More recently, also scholars from several fields have been resorting to Twitter as a professional tool to further their academic careers (Ke et al., 2017; Cooke et al., 2017; Parsons et al., 2014). Social media and Twitter in particular allow researchers to find colleagues with similar interests, initiate collaborations, get inspiration for new research ideas, share new findings, discuss research results and even communicate research with policy makers. Twitter has been increasing in popularity among economists. For instance, many star economists and Nobel prizes have important presence in this social media[1] and participate daily in the #EconTwitter - the sub-network of amateur, professional and academic users who tweet mostly about economics. However, despite the increasing presence of top economists on Twitter and their relative prestige and influence in the academic community, we know little about their activities in the Twittersphere and its relationship with their research outcomes.

In this paper, I aim to fill this gap by investigating how the academic-related activity of the most followed economists on Twitter is related to their research out-

[1]The top 10 list of most followed economists include Nobel prize winners Paul Krugman and Joseph Stiglitz and world renowned economists like Thomas Piketty and Ricardo Haussman.

comes. In particular, I ask if wider research dissemination on Twitter is related to an improvement in scholar's citations and publication metrics. Connecting online patterns of scholarly activity with offline outcomes help to fill the gap in understanding how social media shapes economists careers. The key advantage of focusing on the most influential academic economists for my analysis is that it allows me to assess those economists who serve as inspiration and example for their peers on the social platform and thus are more likely to have their research sharing behavior emulated, which makes them the ideal sample for this study.

In order to identify the social media activities of the most followed economists on Twitter, I use the IDEAS[2] list of economists on Twitter and construct a new dataset merging Twitter-usage data collected using the Twitter API with their scholarly production obtained from the Scopus database. Using this novel database, I systematically study economists' social media activity by constructing a network based on the possible interactions between users on the platform (namely retweets, replies and mentions). Then, I perform a descriptive analysis of the evolution of structure of this interaction network to understand how its features changed over time.

Thereafter, I exploit variation in the number of links to scientific websites shared on the platform to examine the association between Twitter academic activity and research output. The panel nature of my database allows me to examine between-person and within-person effects and to account for time-invariant sources of heterogeneity and overall time trends. Furthermore, I investigate further heterogeneity that may arise from gender differences.

I find that disseminating research on Twitter is correlated to a higher number of citations to an economist in a year, but it is not associated to more publications. I control for the different kinds of twitter activity, namely tweeting in your own Twitter stream, replying other users, mentioning other users or retweeting and find that only science-related activity is associated with the number of citations one receives. The results are robust to a set of individual, field, career time and year fixed effects. These findings suggest that the professional use of Twitter by economists may be associated to the promotion of their academic works without impairing their productivity. While the results shed light on the important connection between online behavior and outcomes, I warn against a causal interpretation of these results because the data I collected and analyzed do not allow for a counterfactual causal reasoning.

Furthermore, I find that the network structure of economists on Twitter is related

[2]IDEAS is the largest bibliographic database dedicated to economics. It is based on Research Papers in Economics (RePEc) database and can be freely accessed.

to the extent that they use the platform as a tool for research dissemination. Unsurprisingly, more influential economists in the network who interact with a broader public share more science-related content. I interpret this as evidence that the most popular economists use Twitter as a research tool aimed to have academic discussion in a public arena. Yet, I find striking differences in Twitter usage between male and female economists. Women tweet less, share less links and, more importantly, share less scientific content. Taken together with the association between sharing scientific links and the accumulation of citations, these results indicate that female economists are not exploiting a potential source of academic visibility.

Given the gender differences in Twitter scientific diffusion, I ask if those disparities are driven by online network differences between men and women. I find evidence that on average women interact with 50% more economists each year and are significantly more influential in the Twitter interaction network. Moreover, women's Twitter networks are more tightly connected than men. This evidence suggests that network differences put women in a prime position to harness the possible benefits of sharing scientific content online.

My paper contributes to a small but growing literature that investigates how academic economists use social media by presenting the first comprehensive analysis of the emergence and evolution of social networks of economists on Twitter (Della Giusta et al., 2021; Khandelwal and Tagat, 2021). Studying scholarly communication online may help to shed light on how economists engage with the broader society and to identify successful strategies of knowledge diffusion (Wehrheim, 2022; Bisbee et al., 2020). Furthermore, connecting the publication dimension with a specific social media activity may shed light on the mechanisms that could be involved in changes of pace and flow of scientific knowledge diffusion inside academia and in the public sphere. My work also contributes to a literature that studies the relationship between altmetric indicators – quantitative and qualitative measures that go beyond the traditional citation-based metrics – and academic outcomes (Bornmann, 2015). To my knowledge, this is also the first study that investigates gender differences in interaction on Twitter between economists. Given its ability to greatly expand the reach of one's research, Twitter could be seen as a tool for equalizing opportunities among researchers. Understanding how gender inequalities reproduces themselves on Twitter also helps us assessing the potential of this social media to help increase the diversity of the academic landscape and help closing the well-documented and persistent gender gap in economics.

The rest of the paper is organized as follows. Section 2 presents the network measures used in this work. Section 3 provides descriptive statistics on academic output, Twitter activities and network interactions, and some stylized facts. Section 4 presents analytical and empirical approach. Section 5 presents the empirical findings. Section 6 concludes the paper.

## 3.2 Background and Conceptual Framework

Twitter is a microblogging platform that allows users to share online snippets of text of at most 280 characters. These short messages may embed images, videos and website addresses (URL). User's tweets are shared in their respective profile timelines ("feed") and can be seen in the general timeline of anyone who "follow" them. There's the possibility of unilateral following, i.e. users are free to follow whomever they want without the need to the following to be reciprocated. Users can then see the tweets of the profiles they follow in their main timeline, which will be shown chronologically or based on the platform's algorithm.

There are three main types of public interaction on Twitter that go beyond following profiles [3]: one can quote and tag someone in their own tweet ("mention"), tweeting commenting or answering a tweet ("reply") or by sharing another user's tweet in your personal feed ("retweet"). Naturally, these interactions can be used to endorse, discuss or criticise other user's tweets. Although each one of these online interactions have it's own particularities, they have one important aspect in common: they require the active interaction by the users, which set them apart from the mostly passive outcomes linked with being a follower. Another interesting aspect of these interactions is that they can be easily tracked over time, which is unfeasible for Twitter followers lists.

The interest in the relationship between academic visibility, scientific paper performance and media exposure is not new: evidence about their association has been documented for more than two decades (e.g. Phillips et al., 1991, Kiernan, 2003). The emergence of the internet and the rise of social media like Twitter have increased the scale, frequency and velocity of spreading information online. As a result, scholarly attention was drawn towards the study of the connection between sharing research on social media and academic outcomes. This interest lead to the surge in popularity of altmetrics (Bornmann, 2015), indicators of research impact and visibility that use non-conventional quantitative and qualitative data to measure dissemination and attention to scholarly work. Tweet counts became a cornerstone statistics of altmetric indicators (Bornmann and Haunschild, 2016; Ortega, 2016).

[3] Twitter also allows for direct messages (DM), which are private and thus not part of the feed of any other user.

Early studies showed a positive correlation between tweet mentions to research papers and citation-metric bases (Eysenbach, 2011; Shuai et al., 2012). However, the observational and uncontrolled nature of these studies made it difficult to disentangle the social media amplification effect from other factors. Subsequent works attempt to disambiguate the Twitter effect on article page views and downloads from other confounding variables performing randomized control trials and failed to find a significant and positive effect of Twitter exposure (Fox et al., 2016; Tonia et al., 2016; Maggio et al., 2019) in biomedical fields and education studies. Nevertheless, the interest on the effects of social media exposure on academic outcomes and diffusion remains high Zhang and Wang, 2021; Klar et al., 2020 and Twitter usage for scientific dissemination is encouraged by several scholars in different fields (Lee, 2019; Cheplygina et al., 2020; Bisbee et al., 2020).

Even if some of the previous literature has included economists on their samples studying scholars on social media, the interest on the use of Twitter specifically by economists, however, is more recent. Since there is some evidence of the relative importance of economists as social scientists on the media and to the broader public, a dedicated would be recommended. For instance, Maher et al., 2020 show that economists have a disproportional presence in congress hearings in the United States from 1946-2016, which may indicate the prestige society and politicians give to the discourse of economists. Wehrheim, 2022 demonstrates how there has been a resurgence of economic experts cited in newspapers in Germany from the 1990s onwards. Given these trends, it is surprising that so little attention has been given to the presence and nature of the activities of economists on such a ubiquitous social as Twitter.

In one of the first analyses dedicated to understanding economists on Twitter, Della Giusta et al., 2021 compare the 25 most followed economists on Twitter with the top 25 scientists in the same social media from other fields. They investigate the differences in social networks, use of language, semantic and sentiment analysis between those two groups and find that economists present a different communication style on Twitter: they use more jargon, use more formal language and engage less with their audience. A closely related study by Khandelwal and Tagat, 2021 collects Twitter usage data from 131 economists working on development studies and combine it with survey data to understand the themes discussed and research dissemination efforts by this community of researchers. They found that most development economists in their sample use social media to share academic research and that their tweets focus on the policy implications of their work in the context of their respective countries. My analysis extends these studies to the investigation of

Twitter usage by economists using a larger sample, longer observation time-window and by connecting their social media activity with their academic performance.

Following the analysis of Della Giusta et al., 2021, to understand how the most popular economists on Twitter interact with each other, I describe a Twitter network as a directed network. Different to their work though, we propose that in this directed network two economists have a link if they mentioned, retweeted or replied the tweet of another economist in a specific year. This interaction network on social media is weighted, i.e. the ties among nodes have weights attached to it, each weight being given by the number of interactions between the two nodes. Based on this interaction network, I calculate a series of centrality measures to understand the structural position of each economist on Twitter. I begin by formally defining a graph that represents networks. A graph comprises a set of nodes $N$ and a $n \times n$ matrix $g$, where $g_{ij}$ represents the relationship between $i$ and $j$. I define five different network measures for each node in the graph to determine their network structure.

*Degree.* Degree is a measure of connectivity that describes how many users an economist interacted with. In a direct graph, degree may be measured in three following ways:

$$\begin{aligned} In - Degree(i) &= \sum_j g_{ji} \\ Out - Degree_i &= \sum_j g_{ij} \\ Degree_i = \sum_j \min\{1, g_{ji} + g_{ij}\} \end{aligned} \tag{3.1}$$

The in-degree captures how many users interacted (mentioned, replied or retweeted) to an economist $i$ on Twitter. The out-degree instead captures how many times the economist $i$ interacted with an user on the platform. Finally, degree is the sum of in-degree and out-degree and captures the overall number of users that have an interaction with economist $i$.

*Closeness Centrality.* Closeness centrality is the reciprocal of the average shortest path between nodes. It captures how "close" a node is to every other node in the network. Closeness centrality is defined as:

$$Closeness_i = \frac{n-1}{\sum_{j=1}^{N-1} d(i,j)} \tag{3.2}$$

where $d(i,j)$ is the shortest path between $i$ and $j$. An economist with lower mean distance to others (higher closeness centrality) have their opinions reaching other Twitter users more quickly than the opinion of an economist with higher mean distance (lower closeness centrality). I normalize this measure so this it is expressed in a scale from 0 to 1.

*Betweenness Centrality.* Betweenness centrality is the sum of the fraction of all-pairs shortest paths that pass through node $i$. Nodes with higher betweenness centrality functions as bridges or brokers between two otherwise relatively disconnected parts of the works. Betweenness centrality is defined as:

$$Betwenness_i = \frac{1}{(N-1)(N-2)} \sum_{s,t \in V} \frac{\sigma(s,t|i)}{\sigma(s,t)} \tag{3.3}$$

where $\sigma(s,t)$ is the number of shortest path between $s$ and $t$ and $\sigma(s,t|i)$ is the number of those path that passes through $i$. The first term in the right-hand side of the equation is a normalization term used in directed graphs, which makes this measure vary between 0 and 1. An economist higher betweenness centrality help connect communities on Twitter not by having many connections themselves but by standing between clusters of interacting economists.

*Clustering Coefficient.* Local clustering coefficient is defined as the fraction of all possible directed triangles that exists through a node $i$. It measures the presence of triadic closures in an ego network. In other words, it measures a probability of a pair of neighbors of $i$ are connected. It is defined by

$$Clustering_i = \frac{2}{Degree(i)(Degree(i)-1) - 2Degree^{\leftrightarrow}(i)} T(i) \tag{3.4}$$

where $T(i)$ is the number of directed triangles through node $i$, $Degree(i)$ is the sum of in-degree and out-degree as defined above in equation 3.1 and $Degree^{\leftrightarrow}(i)$ is the reciprocal degree of $i$. Clustering is undefined when degree is lower than 2 and varies from 0 to 1. The clustering coefficient describe how tightly knit an economist network on Twitter is. High levels of clustering means that users that interact with economist $i$ also interact between themselves. A lower clustering coefficient instead may represent some structure hole, connections that may have been made on Twitter but are not yet established.

## 3.3 Data and Stylized Facts

### 3.3.1 Data Description

I began collecting data on Twitter users in IDEAS RePEc's "Economists on Twitter" list[4]. I focused on the top 25% users with most followers and collected their Twitter handle (i.e. an economist username on the platform) and gender information available in the RePEc database. With this handle I was able to use the Twitter API to collect data on all the tweets they have produced and interacted with, along with the date of creation of the site.[5] Twitter was created in 2006 and the first economist started interacting with each other in 2009. Thus, the observations cover the period 2009-2020. I then calculate the number of tweets of each type (direct tweet, retweets, replies, mentions) and the number of years since the economists created their accounts.

I further elaborated the tweets data by identifying every website URL that was embedded on a tweet text and extracting their domain. A significant share of the URL were shortened, common practice in the platform used to keep the tweets shorter than 280 characters. I processed these shortened URLs and identified all the domains of the website that were shared in the economists' tweets text. To identify if these links had any content that may have academic interest to economists, I collected data from EconLit database, which contain data on the online repositories of all economic journals compiled by the editors of the Journal of Economic Literature. I extract the domain of the journal's online repositories and match them with the ones shared by economists on Twitter.

I then proceed by using the Scopus API to collect data on the publication records of the economists listed in IDEAS Twitter list. I was able to identify publication information of 471 economists. The scholarly production data featured for each economist comprises all scientific documents indexed by Elsevier's Scopus database: papers, conference procedings, books, among other types. I further refine the data sample and focus only on papers published on academic journals. I further complemented this data using the DOI of each paper in the dataset and collecting disciplinary and field of study data from Microsoft Academic Graph (MAG) API. With this data I am able to identify the first publication year and calculate the usual research output measures: number of papers, number of citations and h-index at the paper level and at the individual level.

[4]The updated list can be found at https://ideas.repec.org/i/etwitter.html.

[5]For more information on the Twitter API can be found at https://developer.twitter.com/en/docs/twitter-api/v1/data-dictionary/object-model/user.

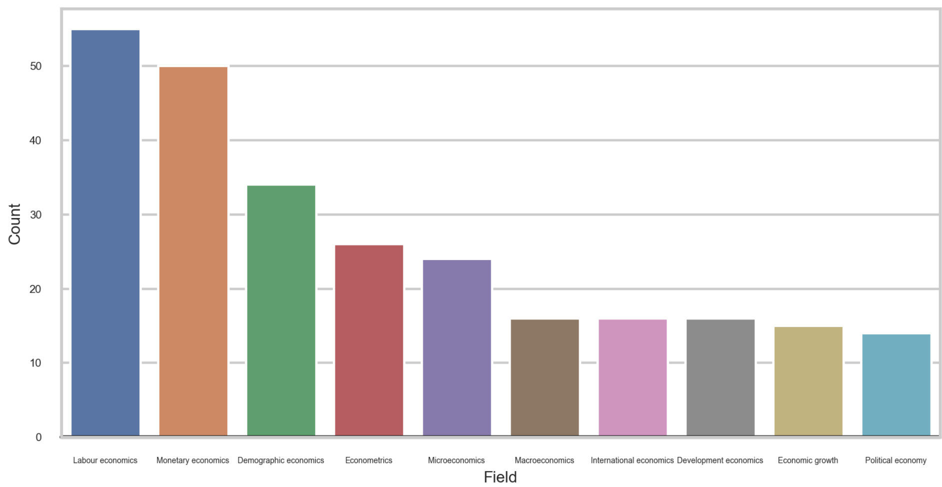
Table 3.1: Descriptive statistics

| | Mean | Median | SD | Min | Max |
|---|---|---|---|---|---|
| ***Panel A - Academic Productivity (N = 471)*** | | | | | |
| Academic Age | 17.75 | 16 | 11.01 | 0 | 54 |
| *Gender* | | | | | |
| Male | 0.84 | 1 | 0.37 | 0 | 1 |
| Female | 0.16 | 0 | 0.37 | 0 | 1 |
| Number of Publications | 2.53 | 2 | 2.67 | 0 | 34 |
| Cumulative Publications | 30.57 | 16 | 43.72 | 0 | 419 |
| Number of Paper | 2.07 | 1 | 2.04 | 0 | 28 |
| Cumulative Paper | 22.62 | 12 | 31.50 | 0 | 335 |
| Number of Citations | 51.65 | 12 | 137.99 | 0 | 1911 |
| Cumulative Citations | 1400.79 | 234 | 3148.24 | 0 | 24385 |
| h-index | 7.91 | 4 | 9.38 | 0 | 70 |
| ***Panel B - Twitter Statistics (N = 451)*** | | | | | |
| Years on Twitter | 10.01 | 10 | 2.21 | 1 | 14 |
| Number of Tweets | 1887.47 | 551 | 3734.32 | 0 | 53687 |
| Number of. Tweets Direct | 1684.98 | 469 | 3457.05 | 0 | 52437 |
| Number of Retweets | 63.09 | 9 | 190.37 | 0 | 3488 |
| Number of Replies | 122.55 | 8 | 354.46 | 0 | 6886 |
| Number of Quotes | 16.84 | 1 | 46.90 | 0 | 915 |
| Number of Links | 808.35 | 226 | 1947.67 | 0 | 62056 |
| Number of Scientific Links | 21.77 | 5 | 48.01 | 0 | 815 |
| Fraction Scientific Links | 0.04 | 0.02 | 0.07 | 0 | 1 |
| Number of Followers | 39130.02 | 9154 | 256803.29 | 0 | 4613276 |
| Number of Following | 1942.54 | 942 | 5132.50 | 0 | 81673 |
| ***Panel C - Network Measures (N = 449)*** | | | | | |
| Degree | 38.97 | 22 | 51.59 | 1 | 549 |
| In-degree | 18.20 | 9 | 27.25 | 0 | 258 |
| Out-degree | 20.77 | 12 | 27.92 | 0 | 291 |
| Degree Centrality | 0.04 | 0.03 | 0.05 | 0 | 0.45 |
| Eigenvector Centrality | 0.03 | 0.01 | 0.05 | 0 | 0.71 |
| Closeness Centrality | 0.24 | 0.26 | 0.09 | 0 | 0.44 |
| Betweeness Centrality | 0.01 | 0 | 0.01 | 0 | 0.19 |
| Clustering | 0.24 | 0.22 | 0.18 | 0 | 1 |

*Notes*: Panel A shows selected measures of productivity of 471 top 25% economists with the most followers on Twitter over the period year 2009 to 2020. Panel B shows descriptive statistics of the 451 economists from those in the top 25% list who were active on Twitter. Panel C shows network statistics of the 449 economists who interacted at least once with each other published by these researchers in the time window 2009-2020.

Table 3.1provides descriptive statistics for research output data, Twitter activity and the network measures calculated for the network described in section 3.2. I draw attention to some distinctive features of the data. Unsurprisingly, the sample is comprised of mostly experienced economists, who have published their first papers in average 17.75 years from 2020. Correspondingly, they are also quite productive: in average they publish 2.53 papers each year and have an h-index of 7.91, which is expected given that their popularity on Twitter probably derives partially from their academic accomplishments. Another important characteristic is that they have in average a long tenure on Twitter, given that the average account age is 10 years. Lastly, the most popular economists on Twitter are mostly males - only 16% of the 471 most popular economists on this social media are women. Despite their intense activity on Twitter (on average they tweet almost 1685 times each year), the number of scientific links they share is quite low - only 21.77 links each year, which represents 4% of the average number of links shared in a year. In terms of networks, economists in average interact with 39 economists in a year and have lower centrality in almost every measure.

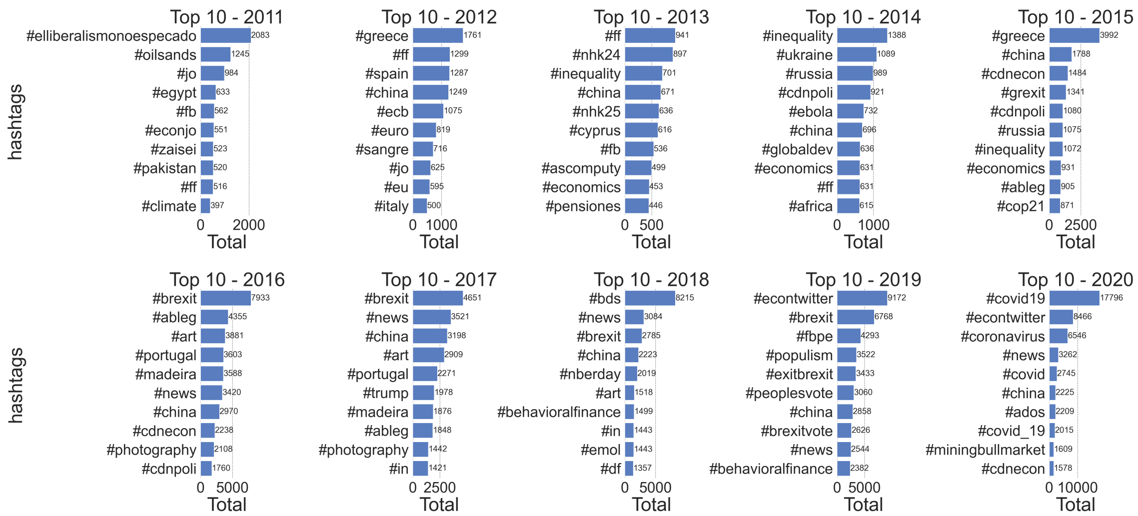
Figure 3.1: Top Fields




This bar plot represents the main field of each economists in our sample based on the categorization given by MAG.

Using the field information of each paper collected with MAG, I identify each economist main area by calculating their modal field, the field that in which he or she published the most. Figure 3.1 describe the 10 most frequent field among the economist on the sample. There is a predominance of labor economists in the sample, who are followed by monetary economists and economists studying demographic themes.

Figure 3.2: Top 10 Hashtags Shared Each Year in the Period 2011-2020

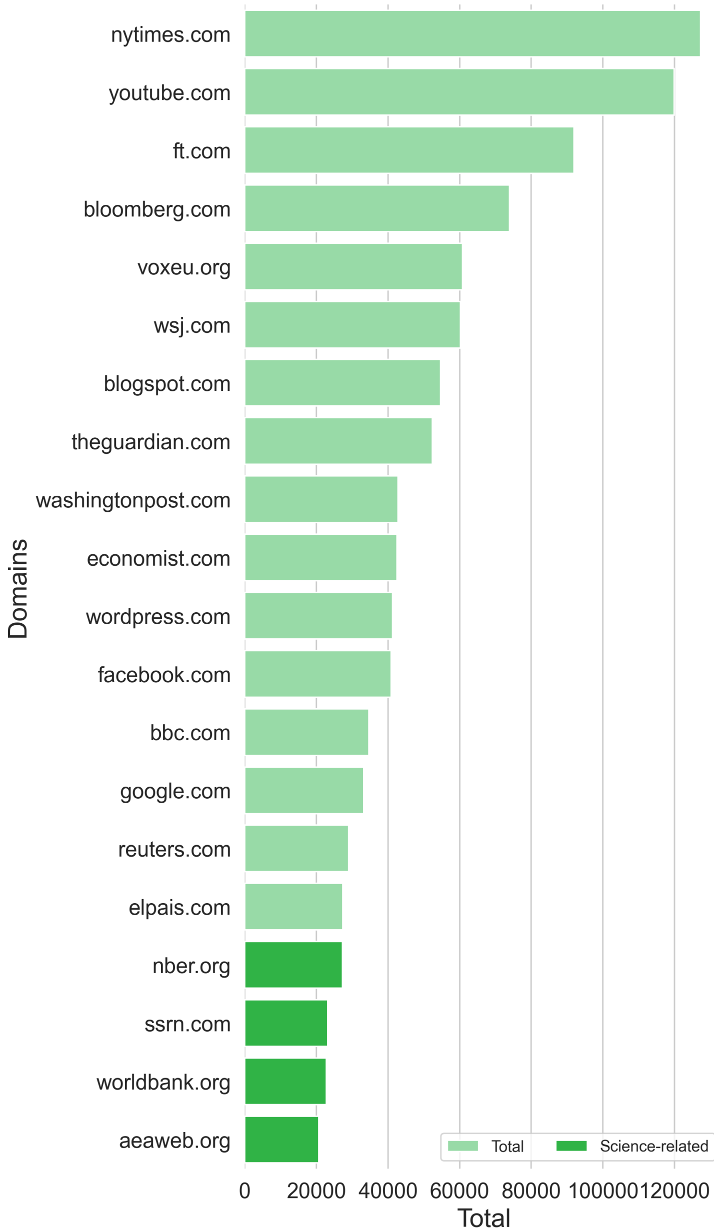



Each horizontal bar plot represents the top 10 most used hashtags on an tweet by economists on the sample in the period 2011-2020. The numbers on each bar represents the absolute number of hashtags shared in a year.

To understand which kind of topics are discussed by the economists on the sample, I analyse the topics they discuss by verifying the distribution of their usage of hashtags. When users prefix a hashtag character in a word or phrase they tag their tweets to a specific topic or subject. This tagging mechanism allows for cross-referencing of content and is used by Twitter to identify trending topics. Several sub-communities identify themselves by the # topic they usually use to share content. One example is the #EconTwitter community described above. Figure 3.2 reports the top 10 most shared hashtags by economists in the sample. As expected, economists trending topics follow the important political and economic events in each year: the Euro crisis, Ukraine protests, Brexit and the COVID-19 pandemics. I draw attention for the emergence of the #EconTwitter community in the last two years as one of the main topics discussed by economists, which signals that research diffusion on Twitter is gaining track among top economists.

Having documented the most popular topics among economists, I turn to examine the website links shared by them. Figure 3.3 shows the top 20 domains shared in the whole 14 years period. It is clear that this list is dominated by news journals from the United States, being the New York Times the most shared source of news shared by economists in the sample. The video platform YouTube is also among the most shared links, which shows the growing importance of video on social media. In the appendix table 3.5 I report the the evolution of the top 10 domains overtime.

Figure 3.3: Top 20 Domains Shared from 2006 to 2020




This figure represents the top 20 domains shared by the sample of economists on Twitter across all the observed period 2006-2020. Darker green bars represent domains linked to repositories of scientific papers.

Despite being relatively few, scientific domain appear in the last positions among the top 20 domains. The most shared links from scientific domains come from the National Bureau of Economic Research (NBER), in which publishes one of the most important working paper series for the economics community. The Social Science Research Network repository comes in second place, followed by the World Bank and American Economic Association (AEA) repositories.

### 3.3.2 Descriptive analysis: Network measures

I start the analysis of the interaction between lead economists on Twitter with a description of the structure of their online network over time. Table 3.2 summarizes the evolution of the Twitter interaction network in the period 2009-2020. Figure 3.6 on the appendix shows the increase of number of economists interacting with each other on Twitter. It also shows that the increase in the number of economists was accompanied by a sharp decrease in the density of the network - the Twitter network is quite sparse. It can be observed that both the number of economists and the mean degree of the network increased over time. It is also worth noting that the transitivity of the network fell in the period while the average path length increased until 2012, when it decreased slightly until 2020 (Figure 3.7 in the appendix). Although the number of components of this network increased considerably from 2009 to 2019, the share of economists who are part of the largest component grew from 12.5% to 76%.

Table 3.2: Measures of Network Structure

| Year | Number of Economists | Mean Degree | Density | Transitivity | Number of Components | Share Largest Component | Average Path Length |
|---|---|---|---|---|---|---|---|
| 2009 | 8 | 1.500 | 0.107 | 0.500 | 8 | 0.125 | 0.000 |
| 2010 | 50 | 3.360 | 0.034 | 0.154 | 30 | 0.220 | 2.436 |
| 2011 | 147 | 6.694 | 0.023 | 0.252 | 61 | 0.517 | 3.095 |
| 2012 | 276 | 8.986 | 0.016 | 0.217 | 93 | 0.601 | 3.492 |
| 2013 | 401 | 11.277 | 0.014 | 0.195 | 125 | 0.676 | 3.480 |
| 2014 | 565 | 12.396 | 0.011 | 0.148 | 187 | 0.669 | 3.317 |
| 2015 | 686 | 14.915 | 0.011 | 0.162 | 220 | 0.673 | 3.123 |
| 2016 | 832 | 18.103 | 0.011 | 0.166 | 238 | 0.715 | 3.045 |
| 2017 | 1027 | 21.190 | 0.010 | 0.168 | 287 | 0.720 | 2.994 |
| 2018 | 1138 | 24.996 | 0.011 | 0.166 | 314 | 0.724 | 2.904 |
| 2019 | 1216 | 27.965 | 0.012 | 0.170 | 283 | 0.765 | 2.858 |
| 2020 | 1251 | 27.260 | 0.011 | 0.170 | 299 | 0.761 | 2.924 |

*Notes*: This table shows the Twitter network structure measures in each year of observation in our sample.

Taken together, these measures indicate that the Twitter interaction network is characterized by being sparse, having a relative low average path length. Although the transitivity has remained stable over the years after 2014, it can be said that there are important signs that the economists' Twitter network shares important characteristics with small-world networks who are characterized by high transitivity and low average path length.

Having describe the structure of the whole Twitter network over time, I turn to examine individual economists' position in this network by comparing their network

Table 3.3: Top 10 economists for each centrality measure.

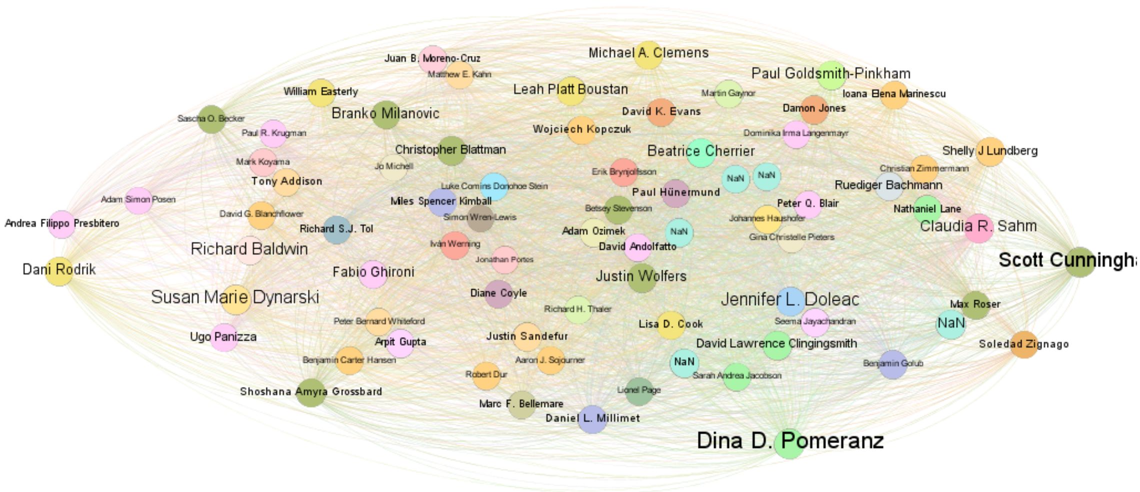
| | Degree Centrality | | | In-degree Centrality | | | Out-degree Centrality | | |
|---|---|---|---|---|---|---|---|---|---|
| Rank | Name | Handle | Mean | Name | Handle | Mean | Name | Handle | Mean |
| 1 | Dina D. Pomeranz | @dinapomeranz | 283.50 | Dina D. Pomeranz | @dinapomeranz | 146.25 | Dina D. Pomeranz | @dinapomeranz | 137.25 |
| 2 | Richard Baldwin | @baldwinre | 237.83 | Justin Wolfers | @justinwolfers | 137.80 | Richard Baldwin | @baldwinre | 123.00 |
| 3 | Leah Platt Boustan | @leah_boustan | 194.20 | Max Roser | @maxcroser | 115.60 | Ruediger Bachmann | @bachmannrudi | 112.50 |
| 4 | Beatrice Cherrier | @undercoverhist | 188.83 | Richard Baldwin | @baldwinre | 114.83 | Claudia R. Sahm | @claudia_sahm | 101.43 |
| 5 | Claudia R. Sahm | @claudia_sahm | 184.43 | Beatrice Cherrier | @undercoverhist | 104.50 | Ugo Panizza | @upanizza | 98.86 |
| 6 | Susan Marie Dynarski | @dynarski | 183.11 | Leah Platt Boustan | @leah_boustan | 102.00 | Susan Marie Dynarski | @dynarski | 98.00 |
| 7 | Ruediger Bachmann | @bachmannrudi | 181.50 | Branko Milanovic | @brankomilan | 89.89 | Scott Cunningham | @causalinf | 93.00 |
| 8 | Justin Wolfers | @justinwolfers | 175.30 | Dani Rodrik | @rodrikdani | 85.50 | Fabio Ghironi | @fabioghironi | 92.83 |
| 9 | Scott Cunningham | @causalinf | 174.00 | Susan Marie Dynarski | @dynarski | 85.11 | Leah Platt Boustan | @leah_boustan | 92.20 |
| 10 | Fabio Ghironi | @fabioghironi | 162.00 | Claudia R. Sahm | @claudia_sahm | 83.00 | Paul Goldsmith-Pinkham | @paulgp | 91.71 |

| | Closeness Centrality | | | Betweenness Centrality | | | Clustering Coefficient | | |
|---|---|---|---|---|---|---|---|---|---|
| Rank | Name | Handle | Mean | Name | Handle | Mean | Name | Handle | Mean |
| 1 | Justin Wolfers | @justinwolfers | 0.40 | Dani Rodrik | @rodrikdani | 0.067 | Eli Dourado | @elidourado | 0.028 |
| 2 | Dina D. Pomeranz | @dinapomeranz | 0.38 | Justin Wolfers | @justinwolfers | 0.063 | Jeremy Horpedahl | @jmhorp | 0.026 |
| 3 | Max Roser | @maxcroser | 0.37 | Richard Baldwin | @baldwinre | 0.057 | Constantin Gurdgiev | @gtcost | 0.026 |
| 4 | Richard Baldwin | @baldwinre | 0.36 | Dina D. Pomeranz | @dinapomeranz | 0.050 | Richard S.J. Tol | @richardtol | 0.024 |
| 5 | Dani Rodrik | @rodrikdani | 0.36 | Branko Milanovic | @brankomilan | 0.048 | Ronan C. Lyons | @ronanlyons | 0.015 |
| 6 | Leah Platt Boustan | @leah_boustan | 0.36 | Michael A. Clemens | @m_clem | 0.042 | Stephen Tapp | @stephen_tapp | 0.014 |
| 7 | Branko Milanovic | @brankomilan | 0.36 | William Easterly | @bill_easterly | 0.037 | Christian Bayer | @christianbaye13 | 0.013 |
| 8 | Beatrice Cherrier | @undercoverhist | 0.36 | Susan Marie Dynarski | @dynarski | 0.037 | Leonardo Becchetti | @leonardobecchet | 0.013 |
| 9 | Alexander Tabarrok | @atabarrok | 0.36 | Diane Coyle | @dianecoyle1859 | 0.036 | Germà Bel | @gebelque | 0.012 |
| 10 | Benjamin Golub | @ben_golub | 0.35 | Tony Addison | @tonysangle | 0.034 | Irvin Rojas | @rojasirvin | 0.012 |

*Notes*: This table shows the rank of economists in the Twitter network for each network centrality measure.

measures. In order to identify the most influential economists on Twitter, I rank the top 10 economist based on each of the centrality measures presented on section 3.2. Table 3.3 report the rankings for the average network measures for each of the 471 most popular economists on Twitter. I draw attention to the important presence of women in the degree rankings. Although the fraction of women in the full sample is 16%, the share of female economists in the top 10 list for degree is around 50% - including the economist that interacted with highest degree, Dina Pomeranz. This pattern do not repeat itself in the rankings for closeness, betweenness and clustering. The economist with higher closeness centrality, betweenness centrality and clustering coefficient are respectively Justin Wolfers, Dani Rodrik and Eli Dourado. As expected, the clustering coefficient ranking comprises of economists that are not on the other lists, given that clustering is often negatively correlated with the other measures in real-world networks.

Figure 3.4: Twitter Interaction Network




This figure represents the interaction network of the economists who replied, mentioned or retweeted the higher number of users in the period 2006-2009. The size of the names and nodes represents economists' degree centrality, while the color of their nodes and links represent their main field of study. Economists who have the same field have same node color.

To illustrate the Twitter interaction network of the most central economists, Figure 3.4 represents the connections over all years between the economists who have degree higher than 250. It can be noted a relative diversity among these economist in terms of gender and field.

## 3.4 Empirical approach

In my empirical analysis I aim to (i) measure the gender disparities in the Twitter citation network; (ii) assess the association between Twitter interaction patterns with sharing scientific papers on the platform; and (iii) relate sharing science with offline research outcomes. I perform a series of regressions examining the variation at the academic publication dimension and Twitter activity.

First, I analyze gender disparities in networks by running regressions of the form:

$$Z_{ifst} = Female_i\beta + X_{it}\delta + \phi_f + \gamma_s + \theta_t + \epsilon_{ifst} \tag{3.5}$$

where $Z_{ifst}$ represents the network measures as defined in section 3.2 for individual $i$ publishing in field $f$, with $s$ years in an academic career observed in year $t$. $Female_i$ an indicator variable for being a woman, $X_{it}$ is a vector of observable individual-year level characteristics including h-index, number of tweets, number of replies, number of mentions and number of retweets. In addition, $\phi_f$ measures field of research fixed effects, $\gamma_s$ controls for academic experience using career-time fixed effects. The parameter of interest in this equation is $\beta$ as I study gender disparities in Twitter network. I add one unit to every count variable (to accommodate observations with zero) and take the logarithm to control for non linearity arising from the skewed nature of their distribution. I use fixed-effects OLS to estimate the parameters.

Then, in order to examine the relation between economists network centrality measures and their sharing of scientific links on Twitter, I run linear regressions of the form:

$$ScienceTweets_{ifst} = Female_i\beta_1 + Z_{it}\beta_2 + X_{it}\delta + \alpha_i + \phi_f + \gamma_s + \theta_t + \epsilon_{ifst} \tag{3.6}$$

where $ScienceTweets_{ifst}$ is the number of links to scientific domains shared by individual $i$, in field $f$, academic age $s$ and year $y$. Additionaly, $Z_{it}$ is a vector of network centrality measures. In this model I include $\alpha_i$, a measures individual fixed effects, to control for unobserved heterogeneity which is constant over time. This means that it is not possible to include time-constant covariates like gender in my estimates. To measure gender differences I use an specification without individual fixed effects and compare the between-individual and within-individual estimates. Our main coefficient of interest in this specification is $\beta_2$, which measures the association between network structure and scientific tweets.

Finally, to study the association between tweeting scientific papers and research output, I run the following regression equation:

$$\begin{aligned} ResearchOutput_{ijft} = & \quad ScienceTweets_{it}\beta_1 + Female_i\beta_2 + Z_{it}\beta 3 + X_{it}\delta + \\ & + \alpha_i + \phi_f + \gamma_s + \theta_t + \epsilon_{ifst} \end{aligned} \tag{3.7}$$

where $ResearchOutput_{ifst}$ denotes the number of papers published and the number of citations accrued in a year by individual $i$, in field $f$, academic age $s$ and year $y$. The coefficient of interest in this model is $\beta_1$ that measures the correlation between sharing scientific articles on Twitter and research output.

## 3.5 Results

In this section I present the empirical results of this chapter. This section is divided in three parts. In the first part I present the estimates of gender disparities in the Twitter citation network. In the second part I present the findings regarding the association between network structure of the economists and their sharing of scientific links on Twitter. Finally, in the last part I show the results of estimates of the association between sharing links online and the offline research output.

### 3.5.1 Gender and Networks

Table 3.4 displays the results from estimations of equation 3.5. I take field, career-time and year fixed effects across all estimates. Standard errors are clustered at the individual level. Column 1 estimates shows that women interact in average with 15.5 more economists than men. The number of tweets and their h-index are positively associated with their degree. Column 2 shows that women are also closer to other economists than men: the average length of the shortest paths linking them to the network is 12.5% shorter (0.03/0.24) than men. Women also have more influence in the flow of information in Twitter: column 4 shows that their betweenness centrality is 20% higher than men (0.002/0.01). In estimates of column 4 and 5, I show the gender differences on clustering. While I do not find a gender gap in clustering, after controlling for degree I find that women network is 9% more clustered than those of men who interact with the same number of economists on Twitter.

In summary, I find significant network differences between men and women on Twitter. This result can be interpreted in different ways. It may be the case that to reach a higher number of followers on Twitter (and consequently entering in the top 25% most popular economists list in which the sample is based), women need

Table 3.4: Gender differences in Twitter Network

| | *Dependent variable:* | | | | | |
|---|---|---|---|---|---|---|
| | Degree | Closeness | Eigenvector | Betweenness | Clustering | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Female | 15.525** | 0.030*** | 0.014*** | 0.002** | 0.015 | 0.023* |
| | (5.133) | (0.004) | (0.002) | (0.001) | (0.010) | (0.010) |
| Degree | | | | | | −0.001*** |
| | | | | | | (0.0001) |
| log(Nb. Tweets) | 14.817*** | 0.021*** | 0.010*** | 0.003*** | −0.014*** | −0.006* |
| | (1.289) | (0.001) | (0.001) | (0.0001) | (0.002) | (0.002) |
| log(h-index) | 7.200*** | 0.013*** | 0.007*** | 0.001*** | 0.002 | 0.006 |
| | (1.894) | (0.002) | (0.001) | (0.0003) | (0.005) | (0.005) |
| Individual Fixed Effects | NO | NO | NO | NO | NO | NO |
| Field Fixed Effects | YES | YES | YES | YES | YES | YES |
| Career-time Fixed Effects | YES | YES | YES | YES | YES | YES |
| Year Fixed Effects | YES | YES | YES | YES | YES | YES |
| Number of Economists | 449 | 449 | 449 | 449 | 449 | 449 |
| Observations | 3,085 | 3,085 | 3,085 | 3,085 | 3,085 | 3,085 |
| $R^2$ | 0.464 | 0.478 | 0.347 | 0.351 | 0.131 | 0.145 |
| Adjusted $R^2$ | 0.441 | 0.456 | 0.319 | 0.323 | 0.094 | 0.109 |

*Notes*: Results estimated using OLS. All regressions include, career-time, year and fields of study fixed effects. Columns 1-5 show the results from estimating gender differences in degree, closeness, eigenvector, betweenness and clustering, respectively. Column 6 show results using clustering as dependent variable and controlling for the degree. Clustered standard errors at the author level in parentheses. Significance levels: † p<0.1; * p<0.05; ** p<0.01; *** p<0.001.

to interact with more users on Twitter than men. Another possible interpretation is that female economists have a higher disposition to interact on social media and act as mediators of knowledge flows in this platform. Given the structure of this data, it is not possible to test these mechanisms. However, these results indicate that female economists are not exploiting Twitter as a potential source of academic visibility as much their male counterparts.

### 3.5.2 Twitter Network and Sharing Science

Table 3.5 report the results of estimates of equation 3.6. First I study the patterns of Twitter usage among economists in the sample. Column 1 show between-individual estimates of network measures correlations and the number of tweets and find that female economists tweet in average 43% less than their male counterparts. I also find that economists with higher degree, closeness and betweenness centrality tweet more, while those with higher clustering tweet less. Perhaps unsurprisingly, economists who are more successful academic careers publish less: academics with 1% higher h-index tweet 37.6% less. The results of the within-individual estimates in column 2 are qualitatively the same: an increase of the degree, closeness and betweenness centrality of an economist is associated with an increase in number of times they

Table 3.5: Twitter Networks and Sharing Science on Twitter

| | *Dependent variable:* | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | log(Nb. Tweets) | | log(Nb. Links) | | log(Nb. Science Links) | | log(Share Science) | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Female | −0.431** | | −0.554*** | | −0.427** | | −0.002 | |
| | (0.137) | | (0.143) | | (0.136) | | (0.003) | |
| Degree | 0.010*** | 0.011*** | 0.009*** | 0.009*** | 0.009*** | 0.007*** | −0.00001 | −0.00003* |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.00002) | (0.00002) |
| Clustering | −0.646** | −0.357** | −0.782** | −0.380** | −0.740*** | −0.521*** | −0.006 | −0.006† |
| | (0.248) | (0.135) | (0.245) | (0.139) | (0.180) | (0.121) | (0.004) | (0.003) |
| Closeness | 4.093*** | 5.688*** | 3.879*** | 5.200*** | 3.090*** | 4.172*** | 0.005 | −0.002 |
| | (0.636) | (0.461) | (0.639) | (0.453) | (0.549) | (0.392) | (0.012) | (0.012) |
| Betweenness | 12.765*** | 12.070*** | 14.016*** | 11.396*** | 15.317*** | 14.386*** | 0.040 | 0.017 |
| | (3.310) | (1.832) | (3.294) | (1.981) | (2.972) | (2.215) | (0.062) | (0.049) |
| log(h-index) | −0.376*** | 0.105† | −0.381*** | 0.034 | −0.215** | 0.028 | 0.001 | 0.0002 |
| | (0.073) | (0.060) | (0.071) | (0.057) | (0.066) | (0.044) | (0.002) | (0.001) |
| Individual Fixed Effects | NO | YES | NO | YES | NO | YES | NO | YES |
| Field Fixed Effects | YES | YES | YES | YES | YES | YES | YES | YES |
| Career-time Fixed Effects | YES | YES | YES | YES | YES | YES | YES | YES |
| Year Fixed Effects | YES | YES | YES | YES | YES | YES | YES | YES |
| Number of Economists | 433 | 449 | 433 | 449 | 433 | 449 | 433 | 449 |
| Observations | 3,085 | 3,189 | 3,085 | 3,189 | 3,085 | 3,189 | 3,085 | 3,189 |
| $R^2$ | 0.461 | 0.844 | 0.503 | 0.855 | 0.459 | 0.796 | 0.145 | 0.600 |
| Adjusted $R^2$ | 0.438 | 0.817 | 0.482 | 0.831 | 0.436 | 0.762 | 0.108 | 0.532 |

*Notes*: This table presents OLS regressions of network measures on different Twitter use statistics, following equation 3.6. The dependent variables are the logarithm of: number of tweets (columns 1-2), the number of tweets with links embedded (columns 3-4), the number of tweets with links to scientific papers (columns 5-6), and the share of scientific links over the total links shared (columns 7-8). All regressions include, field of study, career-time and year fixed effects. Standard errors are clustered at the researcher level. Significance levels: †p<0.1; * p<0.05; ** p<0.01; *** p<0.001.

tweet. However, the number of tweets increases as an economist h-index increases.

Column 3 estimates are analogous to those in column 1 and show the association of the same covariates with the number of website links shared instead of number of tweets. Results are qualitatively the same for number of links, with the difference that the point estimates for the disparities between male and female in sharing links is even higher: 55.5% instead of 43%. Similarly, I find the same qualitative results for the within-individual estimations in column 4 comparing with column 2, with the only difference being that h-index is no longer statistically significant.

Column 5 and 6 displays the estimates of the main results regarding the association between association of network measures and the diffusion of scientific links online. Once again, results are qualitatively the same to the corresponding estimates on column 1 and 4 respectively: higher degree, closeness and betweenness centrality are associated with sharing more scientific links on Twitter while higher clustering is negatively associated to the sharing of links. These results hold between economists and for each individual economist. Lastly, in the estimates of columns 7 and 8 I find that scientific publications shared over the total number of tweets is statistically the same for all economists in the sample. There is some evidence that interacting with more economists in a more clustered network is associated with a lower ratio of scientific papers shared over shared links in the within-individual estimates, however they are small in magnitude.

Taken together, these results indicate that network structure is an important determinant of Twitter activity in general and specifically to the dissemination of research papers online. I also show that network structure is only statistically significant for the extensive margin (i.e. how many scientific links are shared) and not the intensive margin (i.e. the fraction of scientific links over the total links), which may suggest that popular economist found an optimal share of scientific content that they can publish online. Furthermore, I find that women in the sample are less active on Twitter in general. This gender gap may be explained by different mechanisms: maybe women have less available time to use Twitter because they have less leisure time or are more heavily burdened by administrative workload on their non-research work time compared to men.

### 3.5.3 Sharing Science and Research Output

In Table 3.6 I show the association between research diffusion on Twitter and academic performance. In column 1 and 2 I present between-individual and within-individual estimates for the number of papers published in a year, respectively. I find that economists who share more scientific links in average publish more ev-

Table 3.6: Sharing Science and Academic Output

| | *Dependent variables:* | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | log(Nb. Papers) | | | | log(Nb. Citations) | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| log(Nb. Scientific Links) | 0.033** | 0.034** | 0.012 | 0.014 | 0.099* | 0.107* | 0.060† | 0.074† |
| | (0.010) | (0.011) | (0.008) | (0.009) | (0.040) | (0.044) | (0.036) | (0.039) |
| Female | −0.053† | −0.053† | | | −0.025 | −0.030 | | |
| | (0.029) | (0.029) | | | (0.141) | (0.139) | | |
| log(Nb. Tweets Total) | −0.006 | | −0.001 | | −0.077*** | | −0.038† | |
| | (0.006) | | (0.005) | | (0.021) | | (0.022) | |
| log(Nb. Tweets) | | −0.004 | | −0.009 | | −0.082* | | −0.062* |
| | | (0.009) | | (0.007) | | (0.037) | | (0.029) |
| log(Nb. Replies) | | −0.009 | | 0.004 | | 0.001 | | 0.020 |
| | | (0.009) | | (0.007) | | (0.034) | | (0.030) |
| log(Nb. Mentions) | | −0.014† | | −0.009 | | −0.042 | | −0.059* |
| | | (0.008) | | (0.007) | | (0.029) | | (0.026) |
| log(Nb. Retweets) | | 0.013 | | 0.006 | | 0.010 | | 0.017 |
| | | (0.009) | | (0.009) | | (0.040) | | (0.036) |
| log(h-index$_{t-1}$) | 0.146*** | 0.145*** | 0.058** | 0.059** | 0.588*** | 0.587*** | −0.164* | −0.161* |
| | (0.016) | (0.016) | (0.019) | (0.019) | (0.064) | (0.064) | (0.075) | (0.075) |
| Individual Fixed Effects | NO | NO | YES | YES | NO | NO | YES | YES |
| Field Fixed Effects | YES | YES | YES | YES | YES | YES | YES | YES |
| Career-time Fixed Effects | YES | YES | YES | YES | YES | YES | YES | YES |
| Year Fixed Effects | YES | YES | YES | YES | YES | YES | YES | YES |
| Number of Economists | 442 | 442 | 458 | 458 | 442 | 442 | 458 | 458 |
| Observations | 4,154 | 4,154 | 4,306 | 4,306 | 4,154 | 4,154 | 4,306 | 4,306 |
| $R^2$ | 0.253 | 0.255 | 0.567 | 0.568 | 0.280 | 0.281 | 0.562 | 0.563 |
| Adjusted $R^2$ | 0.229 | 0.231 | 0.514 | 0.514 | 0.257 | 0.258 | 0.508 | 0.509 |

*Notes*: This table presents OLS estimates of the equation 3.7. The dependent variables are the logarithm of the number of papers published (columns 1-4) and the number of citations an economists accrued in a 5 years time-window (columns 5-8). All regressions include, field of study, career-time and year fixed effects. Standard errors are clustered at the researcher level. Significance levels: †p<0.1; * p<0.05; ** p<0.01; *** p<0.001.

ery year. The number of tweets is not significant, which means that I do not find differences between economists. Moreover, women in the sample publish 5.3% less papers each year than men. Higher h-index in the last year is associated with a higher number of published papers, as one would expects. The within-individual estimations on columns 3 and 4 suggest that Twitter activity is not associated with economists papers production. None of the coefficients are statistically significant at the conventional levels.

I then turn to the number of citations. Columns 5 and 6 report the estimates of the between-individual estimations and show that economists who share more scientific link accumulate more citations to all their publications than those who share less science. I do not find a significant difference between men and women in terms of number of citations received. Furthermore, a higher number of tweets is associated with a a lower number of citations received in a year. Decomposing the number of tweets by each category, I can see that this association is mainly driven by the number o direct tweets and not by replies, mentions or retweets. In column 7 and 8 I present the within-individual estimates, with show a positive but not highly significant association between the number of scientific links. The results regarding scientific links and the number of tweets of each category are qualitatively similar to those on the between-individual estimates. However, the number of mentions is now significant. The h-index is negatively associated with the number of citations.

In summary, I find that sharing research domains on twitter is positively associated with citations but not with the number of papers published. Several hypotheses can be put forward to explain this association. Given that I do not observe the actual papers being shared, the increase in citation may come from economists using Twitter as a platform to publicize their own papers. Another possible explanation is that being an active science propagator on Twitter may increase one's own visibility and interest on their academic publications in general. It is also interesting to observe that Twitter activity do not observe an association with number of publications, which suggest that the time costs related to Twitter usage does not immediately relates to the knowledge production process, or at least not directly influences it.

## 3.6 Conclusion

In this paper I analyze the online research dissemination activities of the 471 most popular economist on Twitter. By aggregating information from Twitter, Scopus, IDEAS and EconLit dabases, I construct a dynamic network of economists that are connected by their interaction on this social media platform. I document the main topics and website domains shared on Twitter by the most followed economists and evaluate the evolution of the main topological features of their network as well as of individual economists network centrality measures. Interested in the possible gender disparities on Twitter, I document a set of network differences between male and female economists. I then turn to analyze the association between economists' position on the network on the diffusion of research-related links on Twitter. Finally, I study the association between the online diffusion of science-related content and offline research output.

I find that the presence of economists on Twitter has greatly increased in the last 14 years and is still growing. I also find that the network emerging from their interaction on Twitter present features related to small world networks. Furthermore, I document significant disparities between men and women in their interaction in the economist network on Twitter. I also find that the position of economists in the network is a determinant of the amount of research disseminated on the social platform. Taken the result as a whole, I provide what I believe it is the most comprehensive documentation of Twitter usage by economists to date.

This study has several limitations. The sample I observe regard only the most popular economists on Twitter, which diminishes the potential to generalize the results. These concerns are mitigated by the fact that the nature of social media encourages and rewards the emulation of social media strategies. It is also important to reinforce that the results presented in this work are not causal because the nature of the data does not allow us to exploit counterfactual analyzes. Although results cannot be interpreted causally, they show important patterns that shed light on the behavior of scholars online. Moreover, these results provides a further step on the evaluation of the usefulness of Twitter as a research tool and highlight opportunities that can be explored by the community to deal with the lack of diversity of the discipline.

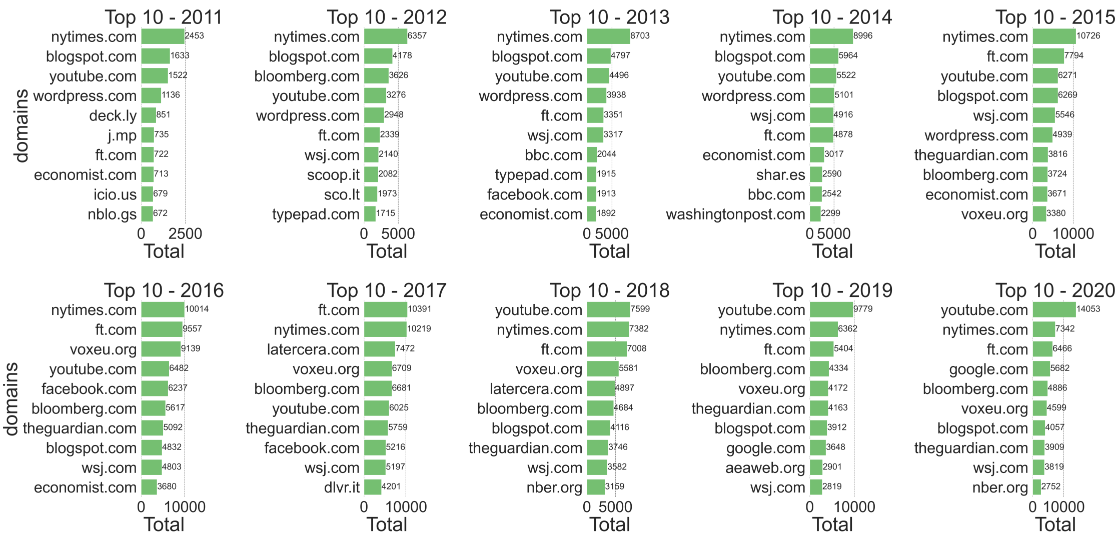
# 3.7 Appendix for Chapter 3

In this appendix we report some additional descriptive statistics and further robustness checks.

## 3.7.1 Additional descriptive statistics

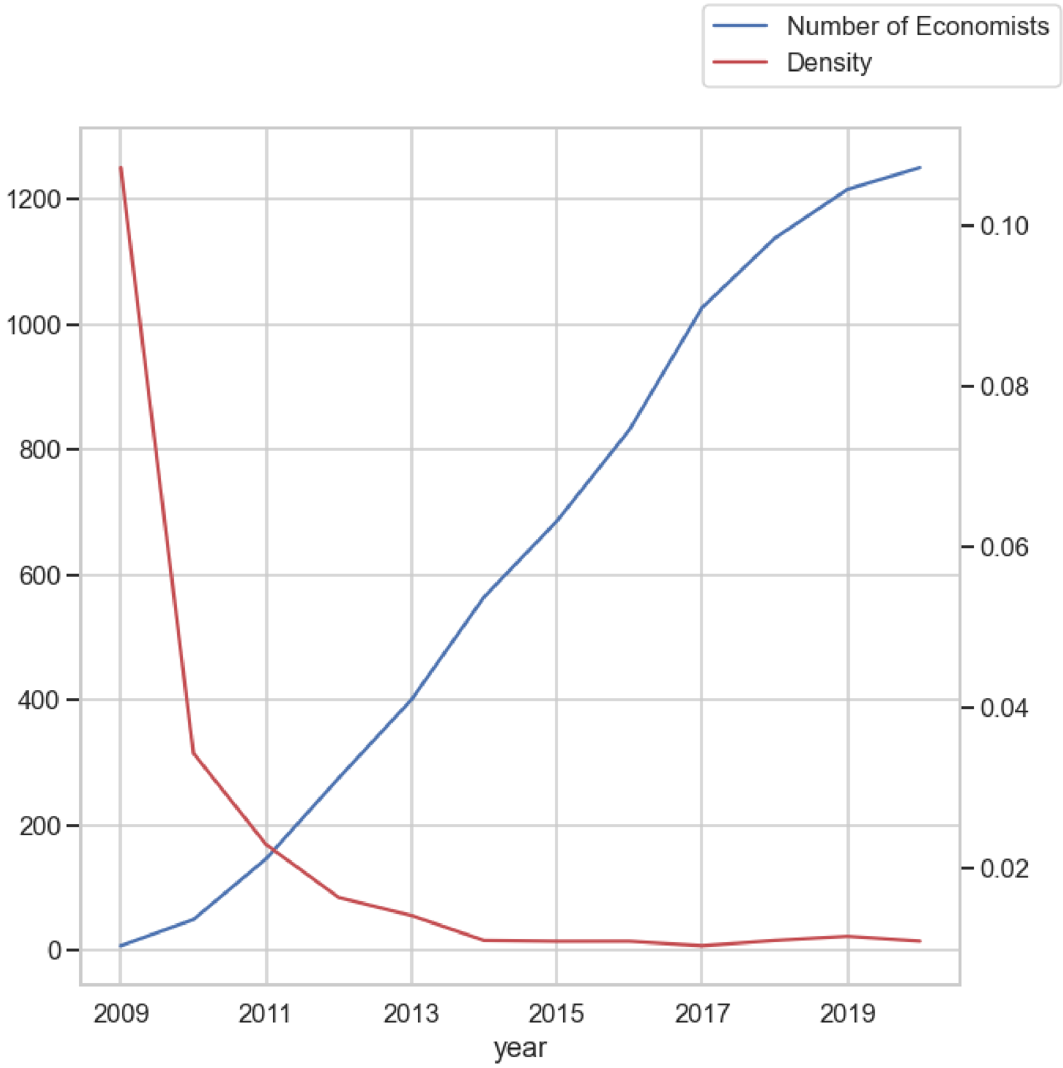
In this section we present additional descriptive statistics of our dataset. In Figure 3.5 I report the top domains shared by economists each year in the period 2011-2020. In Figure 3.6 we describe the evolution of the number of economists and the density of the Twitter Interaction Network. The left vertical axis represents the number of economists and the right vertical axis represents the density of the network. Lastly, in Figure 3.7 I report the evolution of transitivity and average path length of the Twitter Interaction Network over the period 2009 to 2020. The left vertical axis represents the network's transitivity and the right vertical axis represent its average path length.

Figure 3.5: Top 10 Domains Shared Each Year in the Period 2011-2020



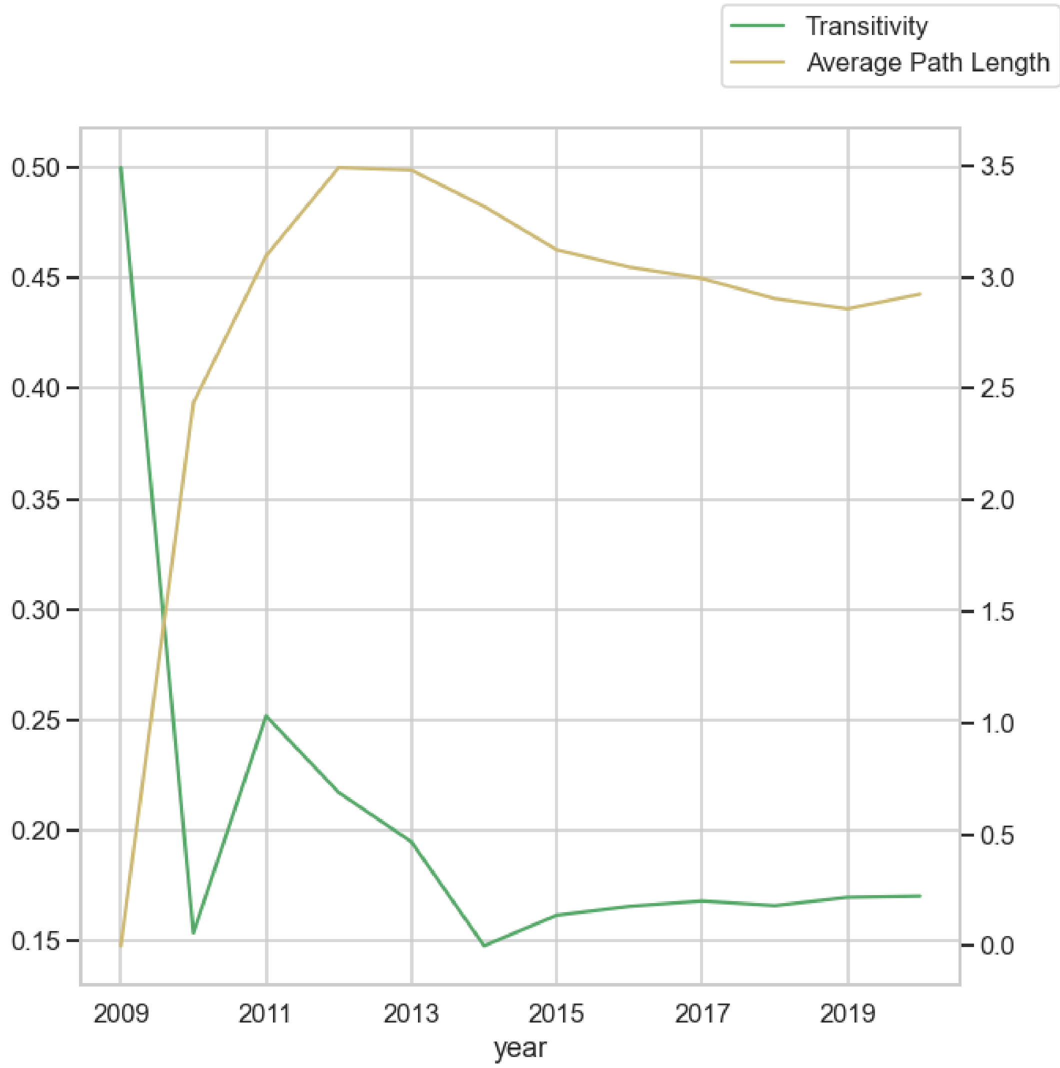

Each horizontal bar plot represents the top 10 most used domains shared on Twitter by economists in the sample in the period 2011-2020. The numbers on each bar represents the absolute number of domains shared in a year.

Figure 3.6: Number of Economists on Twitter and Twitter Citation Network Density




This line plot shows the evolution of the number of economists and the density of the Twitter interaction network in the period 2009-2020. The left vertical axis represents the number of economists and the right axis represents the density.

Figure 3.7: Twitter Citation Network Transitivity and Average Path Length




This line plot shows the evolution of the transitivity and average path length of the Twitter interaction network in the period 2009-2020. The left vertical axis represents the transitivity and the right axis represents the average path length.

# Bibliography


Anderson, K. A. and S. Richards-Shubik (2021). “Collaborative production in science: An empirical analysis of coauthorships in economics.” *The Review of Economics and Statistics*, 1–45. 10.1162/rest_a_01025.

Angrist, J., P. Azoulay, G. Ellison, R. Hill, and S. F. Lu (2020). “Inside job or deep impact? Extramural citations and the influence of economic scholarship.” *Journal of Economic Literature* 58(1), 3–52. 10.1257/jel.20181508.

Arnold, A. et al. (2021). “Perspective: Promoting and fostering multidisciplinary research in universities.” *Research Policy* 50(9), 104334. 10.1016/j.respol.2021.104334.

Battiston, F. et al. (2019). “Taking census of physics.” *Nature Reviews Physics* 1(1), 89–97. 10.1038/s42254-018-0005-3.

Biancani, S., L. Dahlander, D. A. McFarland, and S. Smith (2018). “Superstars in the making? The broad effects of interdisciplinary centers.” *Research Policy* 47(3), 543–557. 10.1016/j.respol.2018.01.014.

Bisbee, J., J. Larson, and K. Munger (2020). “#polisci Twitter: A Descriptive Analysis of how Political Scientists Use Twitter in 2019.” *Perspectives on Politics*, 1–22. 10.1017/S1537592720003643.

Bornmann, L. (2015). “Alternative metrics in scientometrics: a meta-analysis of research into three altmetrics.” *Scientometrics* 103(3), 1123–1144. 10.1007/s11192-015-1565-y.

Bornmann, L. and R. Haunschild (2016). “How to normalize Twitter counts? A first attempt based on journals in the Twitter Index.” *Scientometrics* 107(3), 1405–1422. 10.1007/s11192-016-1893-6.

Bresnahan, T. F. and M. Trajtenberg (1995). “General purpose technologies ‘Engines of growth’?” *Journal of Econometrics* 65(1), 83–108. 10.1016/0304-4076(94)01598-T.

Bromham, L., R. Dinnage, and X. Hua (2016). “Interdisciplinary research has consistently lower funding success.” *Nature* 534, 684–687. 10.1038/nature18315.

Carley, S. and A. L. Porter (2012). “A forward diversity index.” *Scientometrics* 90(2), 407–427. 10.1007/s11192-011-0528-1.

Cedrini, M. and M. Fontana (2018). “Just another niche in the wall? How specialization is changing the face of mainstream economics.” *Cambridge Journal of Economics* 42(2), 427–451. 10.1093/cje/bex003.

Cheplygina, V., F. Hermans, C. Albers, N. Bielczyk, and I. Smeets (2020). “Ten simple rules for getting started on Twitter as a scientist.” *PLOS Computational Biology* 16(2). Ed. by S. Markel, e1007513. 10.1371/journal.pcbi.1007513.

Conti, A. and C. C. Liu (2015). “Bringing the lab back in: Personnel composition and scientific output at the MIT department of biology.” *Research Policy* 44(9), 1633–1644. 10.1016/j.respol.2015.01.001.

Cooke, S. J. et al. (2017). “Considerations for effective science communication.” *FACETS* 2(1). Ed. by V. Metcalf, 233–248. 10.1139/facets-2016-0055.

Della Giusta, M., S. Jaworska, and D. Vukadinović Greetham (2021). “Expert communication on Twitter: Comparing economists’ and scientists’ social networks, topics and communicative styles.” *Public Understanding of Science* 30(1), 75–90. 10.1177/0963662520957252.

Ductor, L., M. Fafchamps, S. Goyal, and M. J. van der Leij (2014). “Social Networks and Research Output.” *Review of Economics and Statistics* 96(5), 936–948. 10.1162/REST_a_00430.

Eysenbach, G. (2011). “Can Tweets Predict Citations? Metrics of Social Impact Based on Twitter and Correlation with Traditional Metrics of Scientific Impact.” *Journal of Medical Internet Research* 13(4), e123. 10.2196/jmir.2012.

Fontana, M., M. Iori, F. Montobbio, and R. Sinatra (2020). “New and atypical combinations: An assessment of novelty and interdisciplinarity.” *Research Policy* 49(7), 104063. 10.1016/j.respol.2020.104063.

Fox, C. S. et al. (2016). “Randomized Controlled Trial of Social Media: Effect of Increased Intensity of the Intervention.” *Journal of the American Heart Association* 5(5), e003088. 10.1161/JAHA.115.003088.

Goel, R. K. and J. R. Faria (2007). “Proliferation of academic journals: Effect on research quantity and quality.” *Metroeconomica* 58(4), 536–549. 10.1111/j.1467-999X.2007.00285.x.

Gruber, M., D. Harhoff, and K. Hoisl (2013). “Knowledge recombination across technological boundaries: Scientists vs. engineers.” *Management Science* 59(4), 837–851.

Guetzkow, J., M. Lamont, and G. Mallard (2004). “What is originality in the humanities and the social sciences?” *American Sociological Review* 69(2), 190–212. 10.1177/000312240406900203.

Hackett, E. J. et al. (2021). "Do synthesis centers synthesize? A semantic analysis of topical diversity in research." *Research policy* 50(1), 104069. 10.1016/j.respol.2020.104069.

Hall, B. H., A. B. Jaffe, and M. Trajtenberg (2001). "The NBER patent citation data file: Lessons, insights and methodological tools." *NBER Working Paper* 8498. 10.3386/w8498.

Hamermesh, D. S. and G. A. Pfann (2012). "Reputation and earnings: The roles of quality and quantity in academe." *Economic Inquiry* 50(1), 1–16. 10.1111/j.1465-7295.2011.00381.x.

Jamali, H., D. Nicholas, and E. Herman (2016). *Emerging reputation mechanisms for scholars: A literature-based theoretical framework of scholarly activities and a state-of-the-art appraisal of the social networking services used by scholars, to build, maintain and showcase their reputations*. Luxemburg: Joint Research Center Publications Office of the European Union. 10.2791/832930.

Jha, Y. and E. W. Welch (2010). "Relational mechanisms governing multifaceted collaborative behavior of academic scientists in six fields of science and engineering." *Research Policy* 39(9), 1174–1184. 10.1016/j.respol.2010.06.003.

Jones, B., E. J. Reedy, and B. A. Weinberg (2014). "Age and scientific genius." *NBER Working Paper* w19866. 10.3386/w19866.

Jones, B. F. (2010). "Age and great invention." *The Review of Economics and Statistics* 92(1), 1–14. 10.1162/rest.2009.11724.

— (2009). "The burden of knowledge and the "Death of the Renaissance Man": Is innovation getting harder?" *The Review of Economic Studies* 76(1), 283–317. 10.1111/j.1467-937X.2008.00531.x.

— (2021). "The rise of research teams: Benefits and costs in economics." *Journal of Economic Perspectives* 35(2), 191–216. 10.1257/jep.35.2.191.

Ke, Q., Y.-Y. Ahn, and C. R. Sugimoto (2017). "A systematic identification and analysis of scientists on Twitter." *PLOS ONE* 12(4), e0175368. 10.1371/journal.pone.0175368.

Khandelwal, A. and A. Tagat (2021). "#DevResearch: Exploring Development Researchers Twitter Use for Research Dissemination." *Scholarly and Research Communication* 12(1). 10.22230/src.2021v12n1a395.

Kiernan, V. (2003). "Diffusion of News about Research." *Science Communication* 25(1), 3–13. 10.1177/1075547003255297.

Klar, S., Y. Krupnikov, J. B. Ryan, K. Searles, and Y. Shmargad (2020). "Using social media to promote academic research: Identifying the benefits of twitter for sharing academic work." *PLOS ONE* 15(4). Ed. by J. Jankowski, e0229446. 10.1371/journal.pone.0229446.

Larsen, P. O. and M. von Ins (2010). “The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index.” *Scientometrics* 84(3), 575–603. 10.1007/s11192-010-0202-z.

Leahey, E. (2016). “From Sole Investigator to Team Scientist: Trends in the Practice and Study of Research Collaboration.” *Annual Review of Sociology* 42(1), 81–100. 10.1146/annurev-soc-081715-074219.

Leahey, E. and S. N. Barringer (2020). “Universities’ commitment to interdisciplinary research: To what end?” *Research Policy* 49(2), 103910. 10.1016/j.respol.2019.103910.

Leahey, E., C. M. Beckman, and T. L. Stanko (2017). “Prominent but Less Productive: The Impact of Interdisciplinarity on Scientists’ Research.” *Administrative Science Quarterly* 62(1), 105–139. 10.1177/0001839216665364.

Lee, J.-S. M. (2019). “How to use Twitter to further your research career.” *Nature*, d41586–019–00535–w. 10.1038/d41586-019-00535-w.

Leone Sciabolazza, V., R. Vacca, T. Kennelly Okraku, and C. McCarty (2017). “Detecting and analyzing research communities in longitudinal scientific networks.” *PloS one* 12(8), e0182516. 10.1371/journal.pone.0182516.

Leone Sciabolazza, V., R. Vacca, and C. McCarty (2020). “Connecting the dots: Implementing and evaluating a network intervention to foster scientific collaboration and productivity.” *Social Networks* 61, 181–195. 10.1016/j.socnet.2019.11.003.

Lindenlaub, I. and A. Prummer (2021). “Network Structure and Performance.” *The Economic Journal* 131(634), 851–898. 10.1093/ej/ueaa072.

Llopis, O., P. D’Este, M. McKelvey, and A. Yegros (2022). “Navigating multiple logics: Legitimacy and the quest for societal impact in science.” *Technovation* 110, 102367. 10.1016/j.technovation.2021.102367.

Maggio, L. A., T. C. Leroux, and A. R. Artino (2019). “To tweet or not to tweet, that is the question: A randomized trial of Twitter effects in medical education.” *PLOS ONE* 14(10). Ed. by L. Lavorgna, e0223992. 10.1371/journal.pone.0223992.

Maher, T. V., C. Seguin, Y. Zhang, and A. P. Davis (2020). “Social scientists’ testimony before Congress in the United States between 1946-2016, trends from a new dataset.” *PLOS ONE* 15(3). Ed. by K.-C. Chang, e0230104. 10.1371/journal.pone.0230104.

Mazzucato, M. (2018). “Mission-oriented innovation policies: challenges and opportunities.” *Industrial and Corporate Change* 27(5), 803–815. 10.1093/icc/dty034.

Microsoft Research (2020). Multi-Sense network representation learning in Microsoft Academic Graph. Microsoft Research. www.microsoft.com/en-us/research/project/academic/artic sense-network-representation-learning-in-microsoft-academic-graph/ (visited on 10/10/2021).

Ortega, J. L. (2016). "To be or not to be on Twitter, and its relationship with the tweeting and citation of research papers." *Scientometrics* 109(2), 1353–1364. 10.1007/s11192-016-2113-0.

Parsons, E., D. Shiffman, E. Darling, N. Spillman, and A. Wright (2014). "How Twitter Literacy Can Benefit Conservation Scientists: Editorial." *Conservation Biology* 28(2), 299–301. 10.1111/cobi.12226.

Phillips, D. P., E. J. Kanter, B. Bednarczyk, and P. L. Tastad (1991). "Importance of the Lay Press in the Transmission of Medical Knowledge to the Scientific Community." *New England Journal of Medicine* 325(16), 1180–1183. 10.1056/NEJM199110173251620.

Porter, A. and I. Rafols (2009). "Is science becoming more interdisciplinary? Measuring and mapping six research fields over time." *Scientometrics* 81(3), 719–745. 10.1007/s11192-008-2197-2.

Rafols, I., L. Leydesdorff, A. O'Hare, P. Nightingale, and A. Stirling (2012). "How journal rankings can suppress interdisciplinary research: A comparison between Innovation Studies and Business & Management." *Research Policy* 41(7), 1262–1282. 10.1016/j.respol.2012.03.015.

Rhoten, D. and A. Parker (2004). "Risks and Rewards of an Interdisciplinary Research Path." *Science* 306(5704), 2046–2046. 10.1126/science.1103628.

Ritzberger, K. (2008). "A ranking of journals in economics and related fields." *German Economic Review* 9(4), 402–430. 10.1111/j.1468-0475.2008.00447.x.

Rylance, R. (2015). "Grant giving: Global funders to focus on interdisciplinarity." *Nature News* 525(7569), 313. 10.1038/525313a.

Shuai, X., A. Pepe, and J. Bollen (2012). "How the Scientific Community Reacts to Newly Submitted Preprints: Article Downloads, Twitter Mentions, and Citations." *PLoS ONE* 7(11). Ed. by C. A. Ouzounis, e47523. 10.1371/journal.pone.0047523.

Singh, J. and L. Fleming (2010). "Lone Inventors as Sources of Breakthroughs: Myth or Reality?" *Management Science* 56(1), 41–56. 10.1287/mnsc.1090.1072.

Sinha, A. et al. (2015). "An overview of Microsoft Academic Service (MAS) and applications." In: *Proceedings of the 24th International Conference on World Wide Web*, 243–246. 10.1145/2740908.2742839.

Smith, T. B., R. Vacca, T. Krenz, and C. McCarty (2021). "Great minds think alike, or do they often differ? Research topic overlap and the formation of scientific teams." *Journal of informetrics* 15(1), 101104. 10.1016/j.joi.2020.101104.

Squicciarini, M., H. Dernis, and C. Criscuolo (2013). "Measuring patent quality: Indicators of technological and economic value." *OECD Science, Technology and Industry Working Papers* 2013/03. 10.1787/18151965.

Sterzi, V., M. Pezzoni, and F. Lissoni (2019). "Patent management by universities: evidence from Italian academic inventions." *Industrial and Corporate Change* 28(2), 309–330. 10.1093/icc/dty070.

Stigler, G. J., S. M. Stigler, and C. Friedland (1995). "The journals of economics." *Journal of Political Economy* 103(2), 331–359. 10.1086/261986.

Stirling, A. (2007). "A general framework for analysing diversity in science, technology and society." *Journal of The Royal Society Interface*. 10.1098/rsif.2007.0213.

Subramanian, A. M., K. Lim, and P.-H. Soh (2013). "When birds of a feather don't flock together: Different scientists and the roles they play in biotech R&D alliances." *Research Policy* 42(3), 595–612. 10.1016/j.respol.2012.12.002.

Sun, Y., G. Livan, A. Ma, and V. Latora (2021). "Interdisciplinary researchers attain better long-term funding performance." *Communications Physics* 4(263). 10.1038/s42005-021-00769-z.

Tartari, V. and S. Breschi (2012). "Set them free: scientists' evaluations of the benefits and costs of university–industry research collaboration." *Industrial and Corporate Change* 21(5), 1117–1147. 10.1093/icc/dts004.

The National Academies (1998). *Trends in the Early Careers of Life Scientists*. Washington, DC: National Academies Press. 10.17226/6244.

— (2005). *Facilitating interdisciplinary research*. Washington, DC: National Academies Press. 10.17226/11153.

Tonia, T., H. Van Oyen, A. Berger, C. Schindler, and N. Künzli (2016). "If I tweet will you cite? The effect of social media exposure of articles on downloads and citations." *International Journal of Public Health* 61(4), 513–520. 10.1007/s00038-016-0831-y.

Trajtenberg, M., R. Henderson, and A. Jaffe (1997). "University Versus Corporate Patents: A Window On The Basicness Of Invention." *Economics of Innovation and New Technology* 5(1), 19–50. 10.1080/10438599700000006.

Twitter (2021). "Number of monetizable daily active Twitter users (mDAU) worldwide from 1st quarter 2017 to 3rd quarter 2021 (in millions)." *Statista*.

Wagner, C. S. et al. (2011). "Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature." *Journal of informetrics* 5(1), 14–26. 10.1016/j.joi.2010.06.004.

Wang, K. et al. (2019). "A review of Microsoft Academic Services for science of science studies." *Frontiers in Big Data* 2. 10.3389/fdata.2019.00045.

Wang, Q. and J. W. Schneider (2020). "Consistency and validity of interdisciplinarity measures." *Quantitative Science Studies* 1(1), 239–263. 10.1162/qss_a_00011.

Wehrheim, L. (2022). "The Sound of Silence: Assessing the (In)Visibility of Economic Experts in German Print Media since the 1960s." *Vierteljahrschrift f??r Sozial- und Wirtschaftsgeschichte* 109(1), 29. 10.25162/vswg-2022-0002.

Wu, L., D. Wang, and J. A. Evans (2019). "Large teams develop and small teams disrupt science and technology." *Nature* 566(7744), 378–382. 10.1038/s41586-019-0941-9.

Wuchty, S., B. F. Jones, and B. Uzzi (2007). "The Increasing Dominance of Teams in Production of Knowledge." *Science* 316(5827), 1036–1039. 10.1126/science.1136099.

Yegros-Yegros, A., I. Rafols, and P. D'Este (2015). "Does interdisciplinary research lead to higher citation impact? The different effect of proximal and distal interdisciplinarity." *PLoS ONE* 10(8), e0135095. 10.1371/journal.pone.0135095.

Zeng, A. et al. (2017). "The science of science: From the perspective of complex systems." *Physics Reports* 714-715, 1–73. 10.1016/j.physrep.2017.10.001.

Zhang, L. and J. Wang (2021). "What affects publications' popularity on Twitter?" *Scientometrics* 126(11), 9185–9198. 10.1007/s11192-021-04152-1.