Collaboration Networks in Economic Science

Thesis

Presented for the Degree of

Doctor of Philosophy

in the School of Economics

University of Cape Town

October 20th, 2018

by

Michael E. Rose

Supervisors:

A/Prof Dr. Co-Pierre Georg

A/Prof David Taylor, PhD

Abstract

When preparing a research article, Economists receive feedback from other academics, present on conference and give talks in seminars. This form of collaboration is termed informal because informal collaborators have, unlike authors, no formal property rights associated with their contribution. However, informal collaboration is so widespread that it appears to be part of the academic production function. Yet, it has received little attention in academia, least in Economics where patterns of informal collaboration differ from that of natural sciences.

Social informal collaboration, the provision of direct feedback, gives rise to a social network. This thesis examines this network. The analysis focuses on the role of individual scientists in the network, which is estimated by different network centralities. Data originate from about 6000 published research articles from six Financial Economics journals between 1997 and 2011.

A theoretical model describes how network centrality proxies the effort informal collaborators exert informally in a project, and how this improves the citation count of the research paper. We then investigate how observable characteristics of authors determine this and other centrality measures and find that common metrics such as productivity and number of citations correlate little with network centrality. As information transmission is an important aspect of social networks we study how network centrality of Economists relates to placement outcomes of their students in the academic job market.

These findings suggest that even informal networks matter in the production of academic research; that these networks contain information above currently used measures of scholarly influence in the profession; and that these networks are used to decrease information asymmetry in the academic labor market.

Dedication

To the scientific mindset.

Acknowledgements

I am grateful to my supervisors, Dr. Co-Pierre Georg and Prof. David Taylor, PhD, for their continuous guidance and insights. Indeed, Co sparked the idea of pursuing a scientific career in my mind, for which I will be forever thankful. Likewise I thank Professor David Taylor, PhD, for great mentorship and a stimulating research environment at the AIFMRM.

Like almost all academic work, this thesis is as well the result of the effort of many minds. In particular, I want to thank Robert Gold (Kiel Institute for the World Economy), Iman van Lelyveld (De Nederlandsche Bank) and Alex Oettl (Georgia Tech, Scheller College of Business) for support, mentorship, advice, insights, training at all stages of my studies. For engaging discussions, exciting and fruitful collaboration and mentorship I thank my other co-authors Daniel Opolot and Suraj Shekhar. Furthermore, I (and my coauthors) thank for insights and comments:

- Pierre Azoulay, Eckhardt Bode, Annamaria Conti, Stan du Plessis, Glenn Ellison, Falko Fecht, Jeffrey Furman, Robert Gold, Ben Golub, Dietmar Harhoff, Edward Kerby, Han Kim, Jaromír Kovářík, Danielle Li, Michael Gofman, Atif Mian, Claire Reicher, Jeff Thursby, Marco van der Leij, Peyton Young and Yves Zenou;
- seminar participants at Oxford University, Kiel Institute for the World Economy, DIW Berlin, University of Cape Town, Georgia Institute of Technology (Scheller), George Mason University, University of Amsterdam, Free University of Amsterdam, Stellenbosch University, Maastricht University, Harvard Kennedy School, Goethe-University Frankfurt and Max Planck Institute for Innovation and Competition;
- IfW/ZBW conference on "The Future of Scholarly Communication in Economics" 2015, the 2015 Biannual Meeting of the Economic Society of South Africa, the HAS workshop "Networks, Complexity & Economic Development" 2015, the "International Symposium on Science of Science“ 2016, the 2016 Africa Meeting of the Econometric Society, the 2016 Annual Meeting of the European Economic Association, 2017 Southern African Finance Association meeting, 'Economics of Scientific Research' workshop at the Erasmus University Rotterdam, and the 17th REER Conference at GeorgiaTech's Scheller College of

Business

For able research assistance, we thank Francis Leni Anguyano, Patricia Chirwa, Aldu Duminy, Samantha Filby, Jesslyn Jonathan, Tiisetso Malinga, Chris McPetrie, Thembekile Ncala, Promise Ngobeni, Khanyisa Yolanda Nogaya and Amiya Sharma.

Thank you to my family and in particular my parents, that were a save haven during extensive travels, a storage for belongings and ideas, and a source of motivation.

Last but not least I acknowledge a scholarship by AIFMRM for the duration of my PhD studies and a Murray-Jelkins travel stipend of UCT's PGFO.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction and Motivation

Understanding research is not only research about research, but also an endeavor into innovation phenomena: How do innovators innovate? What are the elements of the innovation production function? How close should innovators collaborate with others? Do innovators work better alone or in groups? These questions have relevance for academics because collaboration in academia is indeed the norm: Wuchty et al. (2007) document that by the early 2000s most research articles and patents were coauthored by teams, irrespective of field or discipline. At the same time, it takes longer for innovators to innovate (Jones, 2009) and it takes longer for many research articles to be published (Ellison, 2002). The group of innovators I chose to study are academic Economists themselves. While learning about innovation, I wish to contribute to the understanding of the Economics profession itself, where some relevant characteristics are known (Coupé, 2004). The question that I strive to answer in this thesis is: How do Economist's collaborations impact their research and their careers?

Since it studies innovation and collaboration in Economics, this thesis belongs to the small but emerging field of Economics of Science (or Science of Science, or Sociology of Science). Furthering our understanding of the functioning of Science not only guides and informs decisions of individual academics or departments, but also helps organizing science more efficiently, e.g., by improving group formation and talent allocation. The topic of this thesis thus caters to a growing need of funding agencies and the scientific community. Funding agencies are interested to monitor and evaluate the impact of research activities they fund, while the scientific community seeks to understand itself and current inefficiencies in the scientific market. Dedicated funding schemes to study Economics of Science by e.g. Horowitz Foundation, INET, ESRC, CSIC, the European Commission, and NSF give proof of this interest.[1] Finally, there are large-scale gains, too: Improved scientific research and a higher rate of innovation has a direct impact on economic growth (Jaffe, 1989; Adams, 1990; Mansfield, 1991; Cohen et al., 2002).

When individuals interact or when information flows between individuals, there is a social network. In a social network, two individuals are connected and their connections indicate some form of interaction or exchange of information. The same individuals might be connected with other individuals as well, giving rise to complex structures. These structures need to be taken into account: Not only the immediate neighbor's actions or information matter, but also those of individuals somewhere in the network. For example, I might pass on information that I learned from my network neighbor. The appropriate tool to study social interaction and information diffusion is hence social network analysis. Jackson (2014); Jackson et al. (2017) observe that social network analysis has become a popular tool in Economics to study phenomena that involve flow of information and/or social interaction. Social networks are a necessary conduit for the diffusion of information, especially for non-codified (so-called tacit) knowledge, for which there is no other way to diffuse (Singh, 2005). For example, Conley and Udry (2010) and Banerjee et al. (2013) show how social networks impact the diffusion of new opportunities, be it new agricultural technologies or investment decisions. This is exemplified by the anecdotal importance of face-to-face contact in managerial decisions (Storper and Venables, 2004).

Social network analysis not only uncovers information flows among individuals, it also helps understanding spillovers between individuals. Individuals may influence each other in a variety of ways (Jackson and Zenou, 2015). Examples include education, crime, fertility and

[1] Further examples of policies targeted at collaboration networks include the EU-funded Innovative Training Networks and the national Spanish Consolider Program-Ingenio 2010.

market participation (Ioannides, 2012). Social network analysis in the Economics literature borrows from game theoretic insights and from sociological theory. Games on networks start with the basic game theoretic assumption that someone's action's payoff depend on the actions of others. Games on networks generalize those games by considering a diverse set of neighbors: The individuals whose actions I take into account are different from those that these individuals take into account. The classic game theory textbook example, in which two individuals engage in a game, would, if rephrased in network lingo, consider one symmetric link only.

Among the first sociologists to study networks and use social network analysis was Granovetter (1973), who documented how social networks impact job searches. Moreover, he develops a theory according to which individuals usually receive new information from individuals that they rarely meet. This is because network neighbors to which individuals maintain frequent contacts share the same information, simply because they meet often and new sources of information are quickly shaken down. Two large strands of literature then sparked, one that studies social networks in labor markets[2] and one that furthers our understanding of the way social networks diffuse information.

An important question in both the Sociological and Economic literature on social networks, and also in this thesis, is the understanding of the role of particular individuals and their connections. This is a necessary step to foster and shape networks to achieve given ends, since most policies can only target individuals. Various centrality measures help discriminate among individuals Freeman (1978); Friedkin (2015); Ballester et al. (2006). These centrality measures commonly take into account the links between individuals in the entire network, and not just the immediate surrounding of the individual under consideration. As an example, Cruz et al. (2017) recently documented how the centrality of an individual and thus its position in the network predicts that individual's election outcome. Similarly, the network position of the first informed individual predicts the spread of information (Banerjee et al., 2013). It is often "just" predictive power because social networks are inherently endogenous as they may be both cause and consequence of one's action (Graham, 2015).

[2] A more detailed review of this literature is given in chapter 4.

Formally, a social network is a set of individuals, so-called nodes (in graph theory also referred to as vertices), V. Their links, the so-called edges, are denoted by E. The network is then represented by the adjacency matrix $G := (V, E)$. G is a square matrix of dimension $n \times n$ where n is the length of V, or the number of individuals in the graph. G's entries represent links between individual to which row and column indices correspond: $G_{i,j} = 1$ means that nodes i and j are connected. If the graph is unweighted, the entries are binary and take values of 0 or 1, where 1 means that a link exists. If self-links are allowed, the diagonal can take values of 1 as well. If the graph is symmetric, that is, a connection from i to j implies a connection from j to i, the upper triangular mirrors the lower-triangular matrix. In this case the links are said to be undirected. Network links can also differ in weight, e.g. to reflect different intensities of exchange among the linked nodes. In this case the corresponding entry in G can take any non-negative value.[3] Since all information is stored in square matrix G, eigenvalues and corresponding eigenvectors of G are investigated very often. Some eigenvalues of G have an immediate interpretation: The largest eigenvalue is seen as average degree, the second can be interpreted as spectral homophily[4] and the second-smallest eigenvalue (of connected graphs) represents algebraic connectivity. The eigenvector corresponding to the largest eigenvalue contains so-called Eigenvector centralities, which we will meet more often later.

Several authors have investigated social networks in the Economics profession itself. Spurred by the rise in co-authored publications, Goyal et al. (2006) document the corresponding growth of the co-author network for the 1970-1999 period. They find that a small world emerges only by the 1990s. A network with small world properties is one with a high clustering (a high share of network neighbors that know each other), a small average distance and the fact that a large share of authors are somehow connected via intermediate steps. Such a graph would be pictured as dense groups interconnected by few members of the groups: While most authors do not co-author with many different people, most also have someone in their vicinity that does. Small-world networks have unique information transfer capabilities which makes them so special. Understanding the network formation process hence became an emerging topic that still

[3]For some analyses it makes sense to assume negative link weights, but not for this thesis, which is why I refrain from further discussing graphs with negative weights.

[4]Homophily is the tendency of individuals to interact more with each other when they share attributes.

sparks a lot of interest. Theoretical analyses include Li et al. (2013) and Hsieh et al. (2018). This literature borrows heavily from the somewhat older literature on R&D networks of firms.[5] Empirical analyses focus on co-author group formation and try to identify optimal group compositions (Haeussler and Sauermann, 2016; Dahlander and McFarland, 2013).

Parallel to the literature understanding network formation in Economics emerges a literature that estimates the effect of networks on scientifically relevant outcomes. Ductor et al. (2014) show that one's position in the network has predictive power over future research output. Both Azoulay et al. (2010) and Oettl (2012) show that being connected to eminent scholars improves one's own scientific impact, and that this impact consequently diminishes if those eminent scholar pass away. A possible mechanism is that eminent scholars are a source of ideas and conceptual help, that only immediate collaborators have access to. Ductor et al. (2014) bring forward the same explanation for their finding. Colussi (2017) adds to this strand of literature the importance of knowing an editor. This is a network-backed finding that complements Brogaard et al. (2014) who show that authors are more likely to publish in an editor's journal while she is a temporary colleague of the author.

Social networks in science usually mean co-author networks. Few examples have transcended from that and constructed networks based on other characteristics, such as joint faculty positions (Colussi, 2017; Bian et al., 2016). This thesis (in particular in the next two chapters) takes a new step because it infers social networks based on informal collaboration. Informal collaboration refers to the provision of commentary and feedback. It is called informal because – as opposed to coauthorship – there are no formal property rights attached to the production of the research article (Laband and Tollison, 2000). The interesting aspect of informal collaboration is that it likely captures the group of scientists that have contributed to a given research article better than a look at co-authors alone. Ponomariov and Boardman (2016) correspondingly note that often co-authorship is mistaken as the complete research group, which it is not: There are collaborators that provide guidance and are not authors (because authors are those that do the work, which is not necessarily the same as those that have the idea). And there are honorary authors that own critical piece (of lab equipment or data) necessary for the

[5] It will be subject to discussion in chapter 3.

generation of that article.

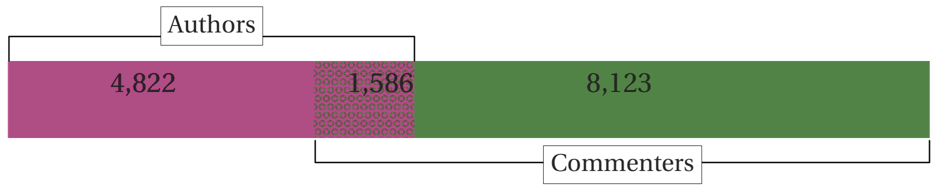
Informal collaboration becomes visible in a special section of the research article upon publication, namely the acknowledgements section. This section is usually positioned as titlepage footnote, sometimes also at the article's end. Cronin (1995) describes and classifies three types of acknowledgments, namely concept-related (ideas, feedback and critique), resource-related (funding, data and materials) and procedure-related (editorial help). For his field of information science he notes that concept-related acknowledgements have been on the rise to the detriment of procedure-related acknowledgements.

This thesis uses hand-collected acknowledgments from 6,401 articles published between 1997 and 2011 in six scholarly journals in finance with a similar focus:[6] The Journal of Finance (JF), the Journal of Financial Economics (JFE), The Review of Financial Studies (RFS), the Journal of Financial Intermediation (JFI), the Journal of Money, Credit, & Banking (JMCB), and the Journal of Banking and Finance (JBF). The first three journals are commonly regarded as the top journals in financial economics. Borokhovich et al. (2000) and annual reports of The Journal of Finance refer to these journals as top journals, too. The vast majority of published research articles in the sample acknowledges informal input by colleagues: Of the 6,401 articles in our dataset, 5,641 ($\approx$ 90%) articles do acknowledge at least one commenter or one seminar or one conference. From each article's acknowledgement section, I collect the number of seminars and conferences, and, crucially, the names of the colleagues that are acknowledged for intellectual input.[7] I focus on concept-related acknowledgments, that Cronin (1995) characterizes as provision of ideas, feedback and commentary. A manual internet-based consolidation procedure for all $19,368$ names in the database is necessary because the same name is frequently spelled in different ways. I am left with 14,531 distinct Economists. Figure 1.1 shows that these are mostly commenters who are not authors (8,123), with the remaining 6,408 being mostly authors who do not appear in any acknowledgement section in our dataset. That is, for every author in our dataset there are 1.3 researchers just commenting on others' work.

[6] The full dataset is available at https://github.com/Michael-E-Rose/CoFE.

[7] The words that authors use to indicate input by their colleagues are usually: comments, insights, encouragements, discussions.

Figure 1.1: Number of authors and commenters in our dataset.



Notes: Graph shows the number of authors of papers and acknowledged commenters in the acknowledgement sections in those papers in our dataset. A commenter is a named researchers mentioned for feedback, advice and discussion (unless she's the journal's managing editor of this or the previous year) in the acknowledgement section of a paper in our dataset. This excludes referees, research assistants and industry professionals, if they are acknowledged as such.

Despite its importance in the academic Economist's daily life, informal collaboration has sparked little research interest. Notable exceptions include Laband and Tollison (2000) and Brown (2005).[8] Laband and Tollison (2000) use 251 featured articles published in the Review of Economics and Statistics during the years 1976-1980. They show a strong positive correlation between the number of acknowledged commenters and the number of citations the article garners. The caliber of a commenter measured in citation count mediates this effect. Brown (2005) tests a sample of 256 research articles published in The Accounting Review, the Journal of Accounting Research, and the Journal of Accounting and Economics during 2000-2002 and show that not only citation count increases, but also acceptance probability at one journal.[9] It is maybe for that reason that Economists are prominently advised to "circulate their papers and give seminars" (Green et al., 2002, p. 1032), to present "work at seminars, professional meetings and conferences" (Hamermesh, 1992, p. 170) and to be "generous" in the acknowledgements section (Thomson, 1999, p. 158).

A major concern with this novel data is that informal collaboration may reflect strategic acknowledging, according to which authors try to influence editors and/or referees by putting down names which they deem helpful to that end. A number of observations speak against the view that the motive to signal quality or steer the editor is the dominant motive, yet the only

[8] Being a pioneer in the study of acknowledgements, Cronin (1995) discusses reasons and implications of what he calls "scholarly courtesy", though not in Economics.

[9] These studies are replicated in chapter 2.4.2.

motive: Names of authors are not sorted by their reputation or prolificness but alphabetically, the list of names often doesn't come first in the acknowledgement section, and it often includes individuals only acknowledged once or twice in the dataset.

Investigating the network of informal collaboration further, chapter 2 starts with interesting descriptives and facts surrounding informal collaboration in financial economics. I show the value of the network derived from informal collaboration in two ways: Variables obtained from the network improve productivity forecasts for focal researchers, above a variables derived from a pure co-author network. The same holds true for explaining academic impact of a research article: Commenter's centrality scores explain a paper's ciation count better than those of the authors, and centrality scores computed in the network of informal collaboration perform better than those compute in a pure co-author network. We document a sharp quality-dispersion within the sample: Publications in higher-ranked journals display display more informal collaboration (comments by colleagues, seminar presentations and conference participations) than those from lower ranked journals.

We then turn to the the individual position of researchers in the network, the so-called centrality, and its determinants. One finding is that citation and publication stock correlate weakly with being acknowledged often. On the other hand, more cited researchers tend to rank higher according to different centrality measures, but less so over time. Most interestingly is however the robust disadvantage that female researchers face. In nearly all years under consideration, female researchers were less often acknowledged on research articles. They were also less central, even for same levels of academic experience and productivity, as their male counterparts, up until the mid-2000s. They continue to be acknowledged less often, but cease to be less central (than comparable males). This leaves a puzzle for future research, and adds to the current debate of females in science (Gaulé and Piacentini, 2017; Chari and Goldsmith-Pinkham, 2017; Wu, 2018; Ductor et al., 2018).

In ranking Economists from the 1997-2011 period, chapter 2 provides a ranking that updates and complements Laband and Tollison (2003) in so far as Economists are not only ranked by the number of times they have been acknowledged, but also by their position in this network.[10] This is contrasted with a ranking based on common metrics such as citations and production of articles. Rankings according to centrality in the network and author metrics hardly overlap, suggesting that the former constitutes a new measure of scholarly influence. This corresponds to the notion of Oettl (2012), who terms often acknowledged individuals helpful.

Chapter 3 lays out a structural equilibrium model where agents' actions are strategic complements (and strategic substitutes) and where knowledge spillovers occur via collaboration. Knowledge spillovers here are modeled as positive complementaries. The model's implication, simply put, is that an academic's return to scientific effort increases if she is connected to individuals with high absorptive capacity. On the project level, this means that informally collaborating with a well-connected commenter increases the impact of that project. Building on the findings of Bramoullé et al. (2014), the model also implies that an informal collaborator's effort brought forward to the project can be approximated in a special centrality measure, the Katz-Bonacich centrality.

We test for this process by hypothesizing that the quality of a research project depends on the Katz-Bonacich centrality of those contributing informally to its development. A major econometric challenge is the endogeneity inherent to all social networks. If not in controlled experiments, researchers employ exogenous variations to the network structure in order to identify network effects.[11] An exogenous variation to network structure is the death of individuals, which has for example been used by Azoulay et al. (2010); Oettl (2012) and Mohnen (2016). We use the exogenous network variation as well and construct two networks, one in which deceased authors are removed and one counterfactual network in which they are still present. The difference in Katz-Bonacich centrality of researchers is then purely due to death, which we use to show that the centrality of informal collaborators matter for the citation count of a research article.

This finding emphasizes the importance of networking and informal collaborating as it shows that even informal networks matter for the production of academic science. It also con-

[10] The complete ranking can be found at `http://www.central-places.net/index`.

[11] A special form of network effects are peer effects, in which characteristics of my network neighbors (the peers) influence mine. Those can be identified exploiting homophiliy and network structure (Bramoullé et al., 2009; Goldsmith-Pinkham and Imbens, 2013; Manski, 1993).

tributes theoretically to the debate on the origin of gains from collaboration. Both Wuchty et al. (2007) and Ductor (2015) show that there is a positive relationship between co-authorship and productivity, besides documenting increased levels of formal intellectual collaboration in academia. According to the model outlined in 3, the positive relationship between intellectual collaboration and research output can be explained by the relationship between information spillover and complementarities in efforts.

Finally chapter 4 turns to the question as to whether social networks matter for scientific careers as well. The particular research question we address is whether students receive better placement outcomes when their adviser is better connected in the Economics co-author network. Identification exploits the longitudinal nature of the dataset, keeping adviser-effects fixed. Those fixed effects capture general helpfulness and maybe other unobserved variables that directly influence her connectedness and student placement simultaneously. Additionally, the regression uses the centrality of the adviser's coauthors as instrument. Combined in a panel, we find that centrality changes positively impact the student's placement. The findings are replicated in the above introduced network of informal collaboration, albeit without explicit identification (for the lack of panel data). The chapter provides supportive evidence to argue that connectedness and improved placement outcomes could result from more central advisers being better suited to disseminate information in the network, which ultimately decreases information asymmetry regarding the student. This evidence results from an investigation of another network, where universities are connected whenever its faculty members coauthor. Exogenous variation due to the death of faculty members or authors somewhere in the network confirm that both link strength between neighboring universities and social proximity between adviser and faculty are associated with a higher placement probability.

These findings are relevant for understanding the student placement process in the job market for academic Economists. The effort dedicated to understanding it can hardly be overstated, because, as Oyer (2006) shows, initial placement matters for an Economists career. The American Economic Association for example devoted an entire ad-hoc committee to the study of this market (Ad Hoc Committee on the Job Market, 2011). Looming open questions include whether new PhDs are allocated efficiently initially and whether market corrections take place.

An interesting puzzle for example is the observation of Smeets et al. (2006), namely that median students of top departments are often worse placed then top students of median universities. Our results provide one possible explanation to this puzzle, namely that advisors from the same university are differently connected.

Furthermore, our results that the connectedness of the adviser matters for the placement of Economics graduates has insights into possible results in the general labor market, too. For example, due to the characteristics of the Economics Job Market, one of which is that there is no information asymmetry regarding job openings, chapter 4 presents some evidence to argue that social networks serve as conduit of information regarding an applicant's quality. While social networks have long been shown to matter in hiring processes (see above), it can be due to two functions of social networks: Providing information of job openings to possible applicants, and providing information on the quality of applicants to possible hirers.

The three chapters use Economics (in particular Financial Economics) as case study. The reasons for this as manifold. For one, I and my coauthors consider ourselves as Economists and believe we have some insights that help us understand the data-generating process better. Secondly, Economics is an interesting field for the study of informal collaboration because there is – unlike in natural sciences – a high importance of informal collaboration (Laband and Tollison, 2000). Natural sciences differ because capital is relatively more important and because the threshold to becoming co-author is lower (as stated in the observation of Haeussler and Sauermann (2016) that co-authors in science publications often don't know each other). This makes Economics an appropriate, if not the most appropriate, field to study informal collaboration. Thirdly, Economics is also regarded by many (not only Economists) as the dominating social science (Fourcade et al., 2015), making it somewhat representative for the other social sciences.

1.2 Organisation of this Thesis

Since my dissertation is by publication, the following three chapters correspond to working papers. The study "What 5,000 Acknowledgements Tell Us About Informal Collaboration in Financial Economics", which is joint work with Co-Pierre Georg, compromises chapter 2. The study "Informal Collaboration with Central Colleagues", which is joint work with Co-Pierre Georg and Daniel Opolot, forms chapter 3. Finally, chapter 4 presents the study "Informal Contacts in Hiring: The Economic Job Market", which is joint work with Suraj Shekhar. Chapter 5 concludes by summarising the findings of the three chapters and discussing ideas for further research.

Chapter 2

What 5,000 Acknowledgements Tell Us About Informal Collaboration in Financial Economics

2.1 Introduction

Collaboration is a crucial ingredient of academic research. Co-authorship has become the norm and leads to research with higher scientific impact, measured by the number of citations research papers receive.[1] Co-authorship is a formal way of collaboration, but it is not the only way in which researchers collaborate. In fact, already Laband and Tollison (2000) highlight that *informal* intellectual collaboration—commenting on a paper or discussing it at a conference or seminar—is prevalent in academic research in Economics.[2] And Colander (1989, p. 146) concludes in his seminal study on research in Economics that *"[i]n studying the Economics Profession, one quickly learns the importance of informal networks, contacts and the exchange of ideas.*

[1]See Beaver and Rosen (1979) for the rise of teamwork in academia and Wuchty et al. (2007) for an analysis of the scientific impact of research produced by teams.

[2]In financial economics, informal collaborations are widespread and occur in the form of verbal and written feedback by other researchers, extended directly, in research seminars or in conferences. Authors are prominently advised by Green et al. (2002, p. 1032) in a joint editorial to *"circulate their papers and give seminars to colleagues to receive constructive criticism before submitting to a journal."*

Much if not most of the debate and discussion about economic ideas take place at the pre-working paper, workshop and working paper stages." The benefits of such informal intellectual collaboration are emphasized by Brogaard et al. (2014), who show that colleagues of editors publish more in the editor's journal during his tenure and that these articles receive significantly more subsequent citations than the average article in the journal. However, this study is an exception and while formal collaboration in Economics has been widely studied (Goyal et al., 2006; Azoulay et al., 2010; Ductor et al., 2014), informal collaboration is much less understood.

In this paper we present the first study of informal collaboration in Financial Economics derived from 6,401 full research papers published between 1997 and 2011.[3] We study Financial Economics because it is both a large and a homogeneous sub-field of Economics. Economics is itself a very good discipline to study informal collaboration because the relation of informal collaboration to formal collaboration tends to be higher for social sciences as compared to natural or life sciences (Laband and Tollison, 2000). We focus on research papers in the "top three" finance journals (the Journal of Finance, the Journal of Financial Economics, and the Review of Financial Studies) and three journals with a lower impact factor (the Journal of Financial Intermediation, the Journal of Money, Credit, and Banking, and the Journal of Banking and Finance). The papers reveal the collaboration patterns of 14,531 researchers, of which only 6,408 are authors. The remaining researchers are acknowledged for helpful input but not having published in our dataset.

We obtain a number of novel results that help us understand collaboration in Financial Economics, and group them along three blocks. The first block studies whether patterns of informal collaboration contain information of interest to economists. We show that this indeed is the case and that a researcher's centrality in the network of informal collaboration correlates with her future productivity and the scientific impact of papers she comments on. This is very useful information for anyone on a hiring committee, in particular when assessing job market candidates who usually do not have a publication yet. The second block of results then studies the network of informal collaboration in more detail. We show that academics who connect disparate research communities are less likely to publish in top finance journals, but if they do,

[3] Of which more than $5,000$ actually acknowledge informal collaboration.

their papers receive an above journal-average number of citations. We also show that many more researchers are involved in the production of research in finance than just the authors of research papers. The consequence of this is that the (social) network of informal collaboration is much more connected than the network of formal collaboration alone, which implies that studying formal collaboration patterns only might miss important aspects of collaboration in finance. Finally, in our third block we combine the researcher and the network perspective by studying what determines a researcher's position in the network of informal collaboration and give a list of the most central researchers in our dataset.

Each of our blocks contains a number of results, to which we now turn one by one. First, we show that the pattern of how a researcher collaborates informally with others contains information about her future productivity, measured as the journal-quality weighted number of future publications. We construct the network of informal intellectual collaboration, where each node is a researcher and each link is an instance of informal intellectual collaboration between two researchers, i.e. one commenting on others' work. We show that including network centrality measures[4] of a researcher computed in the network of informal collaboration, and the researcher's commenters' productivity measures, more than doubles the accuracy of a forecast of the researcher's future productivity, compared to the same measures derived from the network of formal intellectual collaboration.[5]

Second, networks of informal collaboration not only contain information about an author's future productivity, but also about a paper's future academic impact. We find that knowing the average eigenvector centrality of the researchers who have commented on a paper improves the prediction of the future academic impact of the paper over and above the authors' eigenvector centrality–which in itself has predictive power for the paper's future academic impact. Academic impact is measured through the number of citations and whether the paper was published in one of the top three Finance journals. The eigenvector centrality of a researcher

[4]Eigenvector centrality, betweenness centrality, closeness centrality, degree (number of neighbors in the network), and degree of order two (number of neighbors' neighbors in the network).

[5]This analysis follows Ductor (2015), who used a co-author network in Economics. Based on their finding, the authors emphasize that productivity and network position of early collaborators is especially helpful for young researchers, where few other other quality signals are available.

is a well-established measure of influence and power in a network (Bonacich, 1987).[6] Commenters with a high eigenvector centrality are colleagues with a strong influence in the scientific community. Their feedback can thus help authors to better align their paper with what the community considers important and relevant, both in style and content.[7]

Third, authors with a high betweenness centrality in the network of informal collaboration are less likely to publish in one of the top journals in finance, but their articles receive above journal-average citations. Betweenness centrality is an alternative centrality concept which captures how important a researcher is for the flow of information within the network of informal intellectual collaboration (Freeman, 1978). Researchers with high betweenness centrality often connect otherwise disparate research communities. Our result highlights the important role researchers who connect different communities have for the production of research.

Fourth, many more researchers are involved in the production of research in financial economics than "just" the authors of research papers. Some researchers in our dataset provide a lot of commentary and input to other's work, but do not publish a lot themselves. Oettl (2012) calls this trait a researcher's *helpfulness* and distinguishes it from existing measures of performance. He shows that researchers experience a decrease in output quality due to the death of co-authors, and more so if the co-authors were helpful, i.e. often acknowledged. We take this analysis one step further and show that already informal intellectual collaboration with helpful researchers is associated with an increase in the number of citations a paper receives.[8] This is also reflected in the patterns of informal intellectual collaboration of publications in journals with different impact factor: Papers published in the top three finance journals acknowledge substantially more researchers for helpful input, and have been presented at more seminars and conferences.

[6] Banerjee et al. (2013, 2014) for example show how individuals tend to approach well-connected individuals for information sharing and gossip. Ballester et al. (2006) show how effort dedicated in a social network is proportional to someone's eigenvector centrality.

[7] Research papers that receive feedback from more eigenvector central academics are likely to have a higher academic impact than research papers that receive feedback only from colleagues with low eigenvector centrality, all else equal. Identifying the precise mechanism through which feedback from central colleagues helps to improve the academic impact of a paper is beyond the scope of our paper, though. Our goal, instead, is to document this relationship as clearly as possible.

[8] Our setup does not allow us to identify the underlying mechanism, though, so we restrict ourselves to reporting correlations.

Closely related is our fifth finding: The inclusion of networks of informal collaboration dramatically increases the connectivity of the social network. While the network of co-authors is loosely connected and scattered into many different subcomponents, almost all researchers are at least indirectly connected when accounting for informal collaboration. This has implications for the transmission of information and knowledge. For example, small-world networks, which have a high clustering (many researchers work with researchers that also work with each other) and at the same time a small average distance (because some researchers work with others that their collaborators don't work with), emerge more easily than when restricting to formal collaboration only. These small-world networks have unique information diffusion properties and characterize many real-world networks (Watts and Strogatz, 1998; Watts, 1999).

A natural next question then is to ask what determines a researcher's centrality in the network of informal intellectual collaboration. Therefore we, sixth, study what covariates are associated with different centralities. We find that *experience* is negatively associated with having a high eigenvector centrality, controlling for academic prolificness and gender: For every year since the first publication, the researcher is expected to rank 16 ranks worse. By contrast, we find that being very *prolific* (i.e. having a high Euclidean index of citations) is positively associated with betweenness centrality. Being *productive*, i.e. having more publications correlates positively both with eigenvector centrality (about 2 ranks per publication) and betweenness centrality (about 7 ranks per publication). Taken together, these findings show that less experienced researchers with more publications have a more dominant position in the social network of informal intellectual collaboration (eigenvector centrality), but those with more publications are more important for the flow of information within the network, irrespective of their seniority. Finally, we find that female researchers are acknowledged less often in our dataset–even after controlling for academic productivity and experience. Over a three year period, a female reseracher is expected to be acknowledged on 0.3 papers and by 0.5 authors less than male counterparts with similar academic experience and prolificness. Female researchers also have a statistical malus in eigenvector (about 100 ranks) and betweenness centrality (about 120 ranks) less than their equivalent male counterparts, but only until around 2006 and 2005, respectively. This descriptive finding adds to recent debates regarding the role of females in informal collaboration (Chari and Goldsmith-Pinkham, 2017) and collaboration networks (Ductor et al., 2018).

We contrast these findings with those for traditional co-author networks. Here, we find no significant correlation between experience or citation stock with either eigenvector or betweenness centrality. How prolific a researcher is, is mildly negatively correlated with her eigenvector and betweenness centrality. The only variable that significantly and positively correlates with eigenvector and betweenness centrality is a researcher's publication stock, but this correlation is mostly mechanical given how we construct the network. For the network of formal intellectual collaboration, being female has no significant impact on eigenvector or betweenness centrality.

Lastly, we use our dataset to compile a list of the thirty researchers with the highest centralities in the networks of formal and informal intellectual collaboration at different points in time, and include those which have been acknowledged most often. We provide this list to build some intuition about our dataset and also to shine a light on those researchers who have been particularly helpful in the profession.

To date, few studies have investigated the impact of informal collaboration on either the research paper or the author. Cronin (1995) provides a taxonomy of informal collaboration, stating that authors acknowledge individuals for various outsourced tasks (in our study we solely focus on help with conceptual tasks, in Cronin's terminology). Focusing on informal intellectual collaboration in Economics and Biology, Laband and Tollison (2000) find that a higher number of commenters is associated with a higher citation count over seven years. Brown (2005) includes other forms of informal intellectual collaboration, such as seminar presentations and finds that the number of acknowledged seminars is more relevant for citation count than the number of commenters. The same is true for the acceptance probability at prestigious Accounting journals. (To check the external validity of our data, we replicate these studies in section 2.4.2.) Oettl (2012) investigates the impact of informal collaboration on authors in immunology and coins the terms 'helpfulness' for researchers who are often acknowledged on other authors' papers. He finds that losing co-authors with high degree of helpfulness leads to a drop in the quality of a researcher's output by 14%. All three papers point to the relevance of informal intellectual collaboration for productivity.

A growing literature highlights the importance of "knowledge networks" (for a review, see Phelps et al. (2012)). An important variant of these networks are (social) networks of formal collaboration, in which co-authors are linked based on joint publications. The rich literature on co-author networks expands to questions such as (i) how co-author links emerge (McDowell and Melvin, 1983; Freeman and Huang, 2015); (ii) what the individual benefits of network links are to authors (Azoulay et al., 2010; Ductor, 2015); (iii) and whether teams are more productive or influential than solo-authors (Medoff, 2003; Wuchty et al., 2007). But also the topology of the network is of interest, because it affects the speed of learning and the diffusion of information (Alatas et al., 2016). We contribute to this literature by adding the perspective of *in*formal collaboration among researchers which, while less studied, is the more prevalent form of collaboration.

2.2 Data and Variables

2.2.1 Informal Collaboration

To estimate informal collaboration, we manually collect acknowledgments from 6,401 papers published between 1997 and 2011 in six journals in finance with similar focus:[9] The Journal of Finance (JF), the Journal of Financial Economics (JFE), The Review of Financial Studies (RFS), the Journal of Financial Intermediation (JFI), the Journal of Money, Credit, & Banking (JMCB), and the Journal of Banking and Finance (JBF). JF, RFS and JFE are commonly regarded as the top journals in financial economics and the other three journals are comparable in total size.[10] The period was chosen to be in accordance with the coverage of the Scopus database.

From each paper's acknowledgement section or title footnote, we collect the number of

[9] For 4098 of the papers we know the Journal of Economic Literature (JEL) codes from either the published or a previous version. 92% of them belong to general category G (Financial Economics). Additional 6% list E (Macroeconomics and Monetary Economics), but not G.

[10] Borokhovich et al. (2000) and annual reports of The Journal of Finance refer to these journals as top journals, too.

seminars and conferences, and, crucially, the names of the commenters that are acknowledged for intellectual input.[11] We focus on concept-related acknowledgments, that Cronin (1995) characterizes as provision of ideas, feedback and commentary.[12] Like Brown (2005) we omit research assistants, editorial support and non-academic commenters (industry professionals, central bankers) if they are acknowledged as such. If individuals are not thanked for a specific role, we assumed they are acknowledged for concept-related help. From each paper's list of commenters, we also remove the journal's managing editors of the current and the previous two years. This is to avoid a technical overestimation of their importance in the network.[13] In removing editors from the list of acknowledged individuals we also remain comparable to Brown (2005). A manual consolidation procedure for all 19,368 names in our database is necessary because the same name is frequently spelled in different ways.[14] We are left with 14,531 distinct researchers. Figure 1.1 shows that these are mostly commenters who are not authors (8,123), with the remaining 6,408 being mostly authors who do not appear in any acknowledgement section in our dataset. That is, for every author in our dataset there are 1.3 researchers just commenting on others' work.

2.2.2 Researcher Characteristics

In order to compute productivity metrics and the number of current projects, we use information on publication records of both authors and commenters from Elsevier's citation database Scopus. Scopus provides the yearly number of citations for each indexed paper of any author of indexed journal volumes.[15]

A procedure is necessary to link researchers in our dataset with the Scopus database. For

[11] The words that authors use are usually: comments, insights, encouragements, discussions.

[12] Two other forms, which we omit altogether, are resource-related (funding, data and materials) and procedure-related (editorial and moral support).

[13] The vast majority of papers acknowledges the editor of the respective journal. If we calculate an editor's position within the social network of informal collaboration, we are likely to be biased towards more frequently publishing journals. The more paper a journal publishes, the higher is its editor's observed centrality in the uncorrected data.

[14] The Journal of Finance's longtime editor Campbell R. Harvey, for example, is being acknowledged as Cam Harvey, Campbell Harvey, Campbell R. Harvey, and Campell Harvey (with a typo). To avoid wrong aggregations based on typos, we conducted an internet search for every name to obtain the correct one.

[15] Due to Scopus' editorial policy, only select volumes are included. See here for the 2017 Coverage Guide.

researchers that authored a paper in our database we simply use the title of the publication(s) to match author and corresponding Scopus profile. The match of acknowledged commenters who are not also authors in our database follows a more sophisticated procedure because there is no ground truth against which we could evaluate the match. There are two general conditions to match a commenter with a Scopus Author profile: First, the profile is classified by Scopus as working in at least one of the fields "Economics, Econometrics and Finance", "Business, Management and Accounting", and second it does not include more than 2 or 5% of publications in journals outsides these fields. If only one match is found against the Scopus database via a simple name search, we match name and profile. If the search returns less than 5 profiles satisfying above conditions, and they are identical in name and affiliation, we take the profile with the highest publication count. In case more profiles are returned, or the returned profiles do not match in affiliation and/or name, we perform a manual search for all individuals that are acknowledged more than 3 times. As a final quality assessment, we manually look into all profiles that published in an journal where no one else in our database published in, and if necessary correct accordingly.

Following this procedure we match all 6,408 authors and 9,070 out of 11,883 (76.32%) of the acknowledged commenters. If we instead count the number of comments given by researchers that we linked to Scopus and compare it to the total number of comments given, we obtain a coverage of 93.07%.[16] Note that not all acknowledged commenters are represented in the Scopus database: In order to have a Scopus profile, an author must have published at least once in a journal or book that Scopus indexed. Many acknowledged commenters do not satisfy these criteria as they are not academics but industry professionals or research assistants not marked as such. In total we link 11,718 (80.64%) of the researchers in the database to their Scopus profile.

For all researchers in our database which we can map to Scopus, we count the yearly publication stock,[17] and the yearly citation count to these papers.

[16]There are a total of 47,238 comments given or informal collaboration links. 43,966 (or 93.07%) of these are involving an author and a commenter which we could identify on Scopus.

[17]Includes original research articles, books, book chapters and conference proceedings.

As a combination of the two we compute the Euclidean index of citations of an author. The Euclidean index of citations has desirable properties that other indices (such as the h-index) do not possess, and follows the definition of Perry and Reny (2016): For each year t, count the total number of citations to each of i's m publications published until including t, then take the root of the sum of the squared citation counts. More formally, Euclidean index $e_{i,t}$ is

$$e_{i,t} = \sqrt{\sum_{k=1}^{m} c_{k,t}^2}, \tag{2.1}$$

where $c_{k,t}$ is the citation count until t of publication k. An author's Euclidean index of citations hence increases monotonically in the number of publications with positive citation count. Consider for example an author with two publications in t, one which received 5 citations and the other one 50. Her Euclidean index of citations obtains as $\sqrt{5^2 + 50^2} \approx 50.25$. If the first paper receives 5 more citations in $t+1$, and the author publishes another paper that garnered 2 citations, the Euclidean index of citations increases to $\sqrt{10^2 + 50^2 + 2^2} \approx 51.03$.

Using the year of the first publication, we compute the number of years since then for each t between $[1997, \cdots, 2011]$. We call this variable experience. Using the researcher's firstname we estimate her gender using the genderize.io database.[18] We obtain gender estimates for 93% of the 11,718 researchers in our network. Researchers without gender estimate are assumed to be male.

2.2.3 Centrality in the Networks of Intellectual Collaboration

Using acknowledgements of papers and authorship information, we construct two types of networks: In the co-author network (or network of formal collaboration) researchers are connected by an undirected and weighted link whenever they have co-authored a paper in our dataset. Links between two academics are weighted with the number of joint papers. In the network of informal collaboration, two researchers are connected with a weighted and directed link when-

[18] See the genderize.io website.

ever one acknowledges the other as a commenter on a published paper in our dataset. Even though information and spillovers occur in both directions (The commenter provides feedback to the author, and the commenter learns about yet unpublished results to build her own research on), we choose to analyze directed networks because the directionality allows us to easier trace who researchers acknowledge and who they are acknowledged by.

Both networks account for different dimensions of scientific collaboration. A researcher with many links in the co-author network is likely someone who publishes and collaborates often with different individuals. Therefore, the formal network mainly captures productivity. The network of informal collaboration additionally captures a dimension that Oettl (2012) denotes as "helpfulness". This is plausible because a researcher with many outgoing links in the network of informal collaboration represents a frequent commenter. As we show later, a very productive researcher is not necessarily a very helpful researcher in the sense that she is often acknowledged for providing feedback during others' publication process.

Next, we formalize the network construction. For each year t we construct the network using the publications published in that year, as well as in the two previous years, $t-1$ and $t-2$. We construct thirteen networks for all $t = 1999, 2001, \ldots, 2011$, which are all constructed in the same way.[19] Let A_t be the set of papers published in years $\{t, t-1, t-2\}$. To each paper $a \in \{A_t\}$, there is a non-empty set of authors κ_a and a not necessarily non-empty set of commenters ι_a. The resulting network G is weighted in such a way that for each pair (i, j) of academics, g_{ij} increases by $1/|\kappa_a|$. Consider for example a paper by two authors that acknowledged one commenter. We create two directed links between the commenter and each author of weight $1/2$ as well as one undirected (or bidirectional) link with weight 1 connecting the two authors. If one of the authors acknowledges this commenter on another solo-paper, the weight of the link connecting them would increase by 1 to $3/2$. Our weighting scheme thus reflects frequency and intensity of interaction and also corrects for possible misreporting on papers with many authors, i.e. with a large κ_a.

Given the above networks, various centrality measures discriminate among researchers

[19] For this reason we omit the time index when no confusion can arise.

with respect to to their hypothetical access to information traversing the network or according to their possible influence on peers. (Jackson, 2014). Banerjee et al. (2013, 2014) for example show how individuals tend to approach well-connected individuals for information sharing and gossip. The centrality measures we study are degree, betweenness centrality, eigenvector centrality and closeness centrality.

Networks can have different components and, formally, two researchers belong to the same network component if there exists an alternating sequence of researchers and links, called a path, between them. The size of a component is the number of researchers that belong to it. The component containing the most researchers is called the giant component, if at the same time it is also large compared to the rest of the network (Goyal et al., 2006; Jackson, 2014). For technical reasons, we compute all centralities (except degree) in each network's giant component and omit nodes in the other components. This is because the computation of the centralities relies on paths and is hence a component-specific measure. Centralities are not comparable across components: If node i belongs to a small network component, all other nodes are fairly close. In contrast, a node in a large component might have a potentially much smaller centrality because many other nodes are far away.

Degree is a very simple and informative measure in undirected networks, and out- and in-degree in directed networks. Degree takes into account the immediate neighborhood only, leaving aside the global structure of the network. It simply counts the number of links in undirected networks (our co-author networks), or more formally is the size of the neighborhood $N_i(G) = j : g_{ij} > 0$ of researcher i given network G:

$$degree_i = \left|N_i(G_i)\right|, \tag{2.2}$$

Out-degree in the networks of informal collaboration, which are directed (g_{ij} is not necessarily the same as g_{ji}), is the number of unique authors that acknowledge a researcher. In-degree of 1 means that there is only one link directed from the commenter to an author. In-degree is the number of unique commenters an author has acknowledged.

Eigenvector centrality was introduced by Bonacich (1987) and is a measure of influence and power in the network. The vector of Eigenvector centralities of network G is the vector $\mathbf{b}$ satisfying the following relation:

$$\mathbf{b}A = \mu_1(G)\mathbf{b} \tag{2.3}$$

where $\mu_1(G)$ is the leading eigenvalue of G. The eigenvector centrality b_i of researcher i is then

$$\text{eigenvector}_i = \frac{1}{\mu_1(G)} \sum_{j \in N_i} G_{ij} \text{eigenvector}_j \tag{2.4}$$

Eigenvector centrality is hence the weighted count of collaborating researcher, where the weights correspond to their respective eigenvector centralities. Unlike degree, eigenvector centrality captures reach and influence beyond immediate neighbors. Ballester et al. (2006), Hojman and Szeidl (2008) and Elliott and Golub (n.d.) among others show theoretically how an individual's eigenvector centrality is related to equilibrium outcomes in games on networks, as it is directly linked to influence and effort. For these reasons, eigenvector centrality is particularly relevant in provisions of public goods such as knowledge provision, because the effort brought forward in equilibrium corresponds to someone's eigenvector centrality.

Since eigenvector centrality focuses on connectivity and influence only, but remains silent about the importance of a researcher for knowledge flows, we also study betweenness centrality. Betweenness centrality was introduced by Freeman (1978) and is defined as the frequency with which a researcher is on the shortest path between any two researchers (denoted as $\sigma(j,k)$ for researchers j,k) in the network:

$$\text{betweenness}_i = \sum_{j,k \in G} \frac{\sigma(j,k|i)}{\sigma(j,k)} \tag{2.5}$$

Betweenness centrality is often used to measure the individual influence on information flows within a network (Jackson, 2014). A high betweenness central researcher could hold authority over, or control collaboration between, disparate clusters in a network; or indicate they bridge between two otherwise sparsely connected clusters.

In one part of the analysis we use closeness centrality as a measure of a researcher's relative distance to the network (Bavelas, 1950). For n researchers, closeness centrality is the inverse of the average distance of a researcher to other researchers:

$$\text{closeness}_i = \frac{n-1}{\sum_{j\neq i}\sigma(i,j)} \tag{2.6}$$

In order to measure and compare networks in terms of their connectedness, we use network density and average clustering. Density is defined as the share of realized paths $\sum_{i,j}^{G} s_{ij}$ to the number of potential paths $\frac{n(n-1)}{2}$ between a network component with n researchers:

$$\text{density} = \sum_{i,j}^{G} s_{ij}\frac{2}{n(n-1)} \tag{2.7}$$

Density measures the network's efficiency in information transmission. The higher the number, the more potential connections are realized and thus the faster the transmission.

Clustering refers to the connectedness of a researcher's collaborators: How often do a researcher's collaborators collaborate with each other? Formally, a researcher i's clustering coefficient z_i is the share of neighbors that are are connected to each other, over the number of possible pair. For the directed networks of informal collaboration, z_i is defined as:

$$z_i = \frac{|\{e_j k : v_j, v_k \in N_i, e_{jk} \in E\}|}{\text{degree}(i)(\text{degree}(i)-1)}, \tag{2.8}$$

while for the undirected co-author networks it is defined as

$$z_i = \frac{2|\{e_j k : v_j, v_k \in N_i, e_{jk} \in E\}|}{\text{degree}(i)(\text{degree}(i)-1)}. \tag{2.9}$$

For network G with number of researchers n, the average clustering is the sum of all researcher's clustering coefficients divided by their number, n:

$$\text{avg. clustering} = \frac{1}{n}\sum_{j\in G} z_j \tag{2.10}$$

2.3 Informal Intellectual Collaboration in Financial Economics

2.3.1 Improved productivity forecasts

Ductor et al. (2014) show that an economist's future productivity can be forecasted using variables derived from the network of formal collaboration (co-authorship). The productivity of current coauthors of a researcher as well as the researcher's network centrality measures contain information about her future productivity. Future productivity in any given year is the log-transformed journal impact factor-weighted publication count over the next three years. One underlying mechanism may be that researchers become more productive when they have better access to information traversing the network. Such forecasts are relevant for first-time hiring decisions, which in for economics and finance job-market candidates are often based on a limited number of signals, usually the job market paper and maybe a handful of other manuscripts. By looking at coauthors of a job applicant, hiring departments can improve their information about that applicant. Productivity is approximated as weighted publication count, where weights correspond to Journal Impact Factors. Network centralities include degree, second degree, closeness centrality and betweenness centrality, all measured in a co-author network derived from publications in about 100 journals listed in EconLit for a 29-year period.

We follow this method to show that the productivity and network measures derived from a network including informal collaborators (as opposed to just co-authors) increases the prediction power even further. We construct two networks which are not directly comparable to Ductor et al. (2014), but which allow comparison with each other: The first is a co-author network derived from 6,401 research papers published between 1997 and 2011 in our set of six financial economics journals. This network is called "Author network" and links two researchers whenever they have jointly published a paper. The second network, the "Commenter network", is derived from a subset of 5,641 papers, namely all those that include acknowledgements of informal collaboration with other researchers. Two researchers are connected whenever one acknowledges the other on a published paper and follows the construction outlined in the previous section. For each researcher, we compute the following network centralities, network statistics and network neighbor productivity measures: degree, degree of order two, membership in the giant component, closeness centrality, betweenness centrality, the network neighbors' log-transformed sum of individual productivities, and the log-transformed sum of individual productivities of the network neighbors' neighbors.

Table 2.1 presents the result of this exercise. Following Ductor et al. (2014) we use the Root Mean Square Errror (RMSE) as measure of forecast accuracy. A Diebold-Mariano test tests the hypothesis, that a given model and the baseline model are statistically the same. The baseline regression uses observable past output only: (i) cumulative output since the start of a researcher's career until $t-5$, (ii) career time dummies, (iii) a dummy for each year, and (iv) number of years since last publication. The first model then adds a variable measuring recent individual output (the log-transformed weighted count of publications in this and the next two years). This captures a researcher's current projects, which for example are observable to hiring committees, but as a noisy signal since the projects are not yet published papers. Hence this model predicts future productivity if the quality of current projects could be perfectly assessed. The third model includes variables derived from the co-author network. It increases prediction accuracy by 4.2% over the baseline model. The fourth model uses variables derived from the network of informal collaboration instead. Its prediction accuracy increases by more than 8.3%. Combining variables derived from both networks increases prediction accuracy by more than 9.7% over the baseline.

We use the exact same variable definitions as Ductor et al. (2014) but differ in sample size. While Ductor et al. (2014) use publications over 29 years from journals listed in the EconLit database, we include 6,401 publications from 6 journals over 15 years. Our measure of productivity however includes all publications listed on Scopus (weighted with the same Journal

Table 2.1: Comparison of forecasts of researcher productivity akin to Ductor et al. (2014).

	Adj. R^2	RMSE	RMSE Differential
Benchmark	0.17	1.43	
Recent past output	0.57	1.03	27.97***
Author network variables	0.23	1.37	4.20***
Commenter network variables	0.30	1.31	8.39***
Auth. net. and com. net. variables	0.32	1.29	9.79***
All	0.57	1.02	28.67***

Notes: ***, ** and * indicate statistical significance to the 1, 5 and 10 percent level in Diebold-Mariano tests, respectively. Table compares different forecasts for academic productivity to a benchmark forecast, according to Ductor et al. (2014). RMSE is the root-mean-square error of the corresponding regression. "RMSE Differential" is the difference in the RMSE over the baseline. Statistical significance levels correspond to a which a Diebold-Mariano test, whose Null Hypothesis is that the model is the same as the benchmark model. Variable construction corresponds to Ductor et al. (2014) but uses our set of papers derived from six financial economics journals over a 15-year period.

Impact Factors as used by Ductor et al. (2014), if available), not just those listed in EconLit. As more publications are captured, productivity measures are thus possibly more complete.

2.3.2 Predicting a paper's academic success

Laband and Tollison (2000) and Brown (2005) both relate a paper's academic impact to the amount of informal collaboration. Laband and Tollison (2000) show that a paper's citation count correlates strongly with the number of acknowledged commenters and the commenters' past academic productivity. They study a sample of 251 feature articles from The Review of Economists and Statistics. Brown (2005) shows that acceptance probability increases with the number of acknowledged conferences, while controlling for the number of acknowledged seminars and commenters. He studies a sample of 305 papers submitted to The Accounting Review. In section 2.4.2 we show that these observations hold for Financial Economics too, using a much larger sample. Starting from these results, we show that the centrality of authors and acknowledged commenters in the network of informal intellectual collaboration contains information about the paper's scientific impact, over and above the measures of the author's and commenters' past academic productivity. As networks are a necessary conduit for information, more central commenters are better able to grasp information, which is potentially critical for ongoing or future research projects. Interestingly, the author's and commenters' centrality in the network of informal intellectual collaboration correlates stronger with the paper's academic success than the centrality in the network of formal collaboration (the co-author network). A plausible reason for the higher information content of networks of informal collaboration is that they capture a researcher's connectedness – and hence ability to receive traversing information – better than co-author networks.

We model the success of paper p published in year t as:

$$\text{Success}_p = \text{Characteristics}_p + \text{Author centrality}_{p,t-1} + \text{Commenter centrality}_{p,t-1} + \mathbf{D}_t + \varepsilon_t, \tag{2.11}$$

where we measure Success in two ways, as the citation count of paper p according to Scopus in July 2018[20] and as an indicator variable that equals one if the paper was published in a top three finance journal and zero otherwise. Characteristics is a vector with the following controls per paper p: Number of pages, number of authors, number of acknowledged commenters, the author's total Euclidean index of citations according to equation (2.1) in the year before publication, and the commenter's total Euclidean index of citations in the year before publication. When citation count is the dependent variable, we also include a fixed effect for each journal to pick up journal-specific effects such as popularity, topic, and quality of the editorial process. *Author position* is either the authors' sum of betweenness centralities (equation (2.5)) or sum of eigenvector centralities (equation (2.4)). We first use centralities derived from the network of formal collaboration corresponding to $t-1$, and compare them to centralities derived from the network of informal collaboration corresponding to $t-1$. We use the network in the year before publication to avoid that the links observed from the paper itself in some way influence the network position of either authors or commenters. In either case, centralities are computed in the network's largest component. Finally we add fixed effects for the publication year, to e.g. account for a different number of years an article could garner citations.[21]

[20] Using citations according to Web of Science do not materially change our results.

[21] Using the paper's citation count over the next six years and removing journal-fixed effects leaves the regression results almost unchanged.

We study two samples. The first sample uses the co-author network's giant component and consists of 363 papers where the author(s) and at least one author is in the previous year's network. The second sample uses the commenter network and consists of $3,432$ papers where the author(s) and at least one of the acknowledged commenters has also been acknowledged in the previous three years (which is a requirement to be member in the network of informal collaboration for the year before publication).

Tables 2.2 and 2.3 report estimation results of the model above, using centralities computed in the largest component of the co-author network and the commenter network, respectively. Models (1) and (2) are marginal effects of a negative binomial regression with citation count in July 2018 as the dependent variable. Models (3) and (4) are marginal effects of a logistic regression model with top journal publication dummy as the dependent variable. Marginal effects give the expected percentage increase in the number of citations if the explanatory variable was to increase by 1 unit, holding all other variables constant at their mean and setting binary variables equal to 0.

There are three observations we would like to highlight based on these correlations. First, the sum of commenters' network centrality contains information for both citation count and journal publication probability, above the information embedded in the aggregated network centralities of authors. This can be seen from the fact that coefficients for commenters' centralities are statistically significant in all models. In some cases commenters' network positions are statistically significantly correlated with our dependent variables, while the network positions of co-authors are not.

Second, and most importantly, models with centralities computed in the commenter network outperform those models with centralities computed in the co-author network. Direct comparison of tables 2.2 and 2.3 shows that measures of goodness-of-fit (Akaike Information Criterion for models (1) and (2), and R^2 for models (3) and (4)) indicate better fit, albeit the difference is small in some cases. The exception is model (4) which correlates eigenvector centrality of commenters with top journal publication. This model explains about 50% more variance when computed in the commenter network as compared to the co-author network.

Finally, the centrality of authors and commenters matter in crucial ways. Specifically, betweenness centrality and eigenvector centrality of authors and commenters in both networks are correlated in different ways with the dependent variables. Comparing model (2) of tables 2.2 and 2.3, we see for example that the commenters' sum of eigenvector centralities (ability to influence the network) in the co-author network explains statistically significantly the paper's academic impact, but only when computed in the commenter network. The effect is economically significant, too. If the commenter's total eigenvector centrality increases by one standard deviation, the average paper's citation count is expected to increase by roughly four citations. Comparison of models (1) and (3) of table 2.3 shows that the betweenness centrality in the commenter network of both authors matters positively for citation count, but is associated negatively with top journal publication. On the other hand, eigenvector centrality is positively associated with publishing in a top three finance journal and with citation count. Recalling the definitions of the two centralities, this finding suggests that authors and commenters that connect different communities (high betweenness centrality) are less likely to publish in one of the top journals but once published, their papers are cited more than the average of the journal. Authors and commenters that are well connected in the community and better positioned to exert influence (high eigenvector centrality) are more likely to publish in one of Finance's top journals and also to get cited more than the journal's average.

Table 2.2: Results of citation count and journal publication correlation using network measures of the co-author network.

	Total citation count		Top publication	
	negative binomial		*logistic*	
	(1)	(2)	(3)	(4)
Auth. total betweenness	-0.110		-0.274	
	$p = 0.696$		$p = 0.811$	
Com. total betweenness	-0.159		1.638^{*}	
	$p = 0.381$		$p = 0.082$	
Auth. total eigenvector		0.734^{**}		-1.992^{*}
		$p = 0.027$		$p = 0.095$
Com. total eigenvector		0.382		-1.265
		$p = 0.129$		$p = 0.160$
Constant	4.192^{***}	4.261^{***}		
	$p = 0.000$	$p = 0.000$		
Means	91	91	0.501	0.501
Article characteristics	*Yes*	*Yes*	*Yes*	*Yes*
Journal-fixed effects	*Yes*	*Yes*	*No*	*No*
Publication year-fixed effects	*Yes*	*Yes*	*Yes*	*Yes*
N	363	363	363	363
R^2			0.347	0.357
Akaike Inf. Crit.	3,955.394	3,945.847		

Notes: ***, ** and * indicate statistical significance to the 1, 5 and 10 percent level, respectively. *Total citation count* is the paper's sum of citations until July 2018. *Top publication* equals 1 if the paper was published in either the JF, RFS or the JFE, and 0 otherwise. *Auth. total betweenness* and *Com. total betweenness* is the authors' resp. acknowledged commenters' combined betweenness centrality (equation (2.5)). *Auth. total eigenvector* and *Com. total eigenvector* is the authors' resp. acknowledged commenters' combined eigenvector centrality (equation (2.4)). All values are measured in the giant component of the co-author network corresponding to the year before publication. *Article characteristics* includes the number of pages, the number of authors, the number of acknowledged commenters, the authors' combined Euclidean index of citations and the commenter's combined Euclidean index of citations (equation (2.1), both of which are measured in the year before publication using data from Scopus. Only papers published in six financial economics journals published between 1997 and 2011 considered considered where the author(s) and at least one of the acknowledged commenters (excludes the journal's managing editor) is part of the giant component of the co-author network in the year before publication.

Table 2.3: Results of citation count and journal publication correlation using network measures of the commenter network.

	Total citation count		Top publication	
	negative binomial		*logistic*	
	(1)	(2)	(3)	(4)
Auth. total betweenness	8.380^{***}		-10.055^{**}	
	$p = 0.000$		$p = 0.036$	
Com. total betweenness	1.882^{***}		11.252^{***}	
	$p = 0.0003$		$p = 0.00000$	
Auth. total eigenvector		1.137^{***}		13.602^{***}
		$p = 0.002$		$p = 0.000$
Com. total eigenvector		0.313^{**}		9.214^{***}
		$p = 0.015$		$p = 0.000$
Constant	4.012^{***}	3.934^{***}		
	$p = 0.000$	$p = 0.000$		
Means	91	91	0.501	0.501
Article characteristics	*Yes*	*Yes*	*Yes*	*Yes*
Journal-fixed effects	*Yes*	*Yes*	*No*	*No*
Publication year-fixed effects	*Yes*	*Yes*	*Yes*	*Yes*
N	3,432	3,432	3,432	3,432
R^2			0.468	0.544
Akaike Inf. Crit.	37,071.510	37,103.170		

Notes: ***, ** and * indicate statistical significance to the 1, 5 and 10 percent level, respectively. *Total citation count* is the paper's sum of citations until July 2018. *Top publication* equals 1 if the paper was published in either the JF, RFS or the JFE, and 0 otherwise. *Auth. total betweenness* and *Com. total betweenness* is the authors' resp. acknowledged commenters' combined betweenness centrality (equation (2.5)). *Auth. total eigenvector* and *Com. total eigenvector* is the authors' resp. acknowledged commenters' combined eigenvector centrality (equation (2.4)). All values are measured in the giant component of the commenter network corresponding to the year before publication. *Article characteristics* includes the number of pages, the number of authors, the number of acknowledged commenters, the authors' combined Euclidean index of citations and the commenter's combined Euclidean index of citations (equation (2.1), both of which are measured in the year before publication using data from Scopus. Only papers published in six financial economics journals published between 1997 and 2011 considered where the author(s) and at least one of the acknowledged commenters (excludes the journal's managing editor) is part of the giant component of the commenter network in the year before publication.

2.3.3 Social Networks of Informal Intellectual Collaboration

Having investigated how centrality in the network of informal collaboration relates to academic outcomes, we now turn to an analysis of the network structure itself. Tables 2.4 and 2.5 present basic statistics regarding the two network types. There are several things we want to point out with these tables. First, as the number of publications grows, the number of researchers (authors, commenters) grows too, and at about the same pace. The 1999 network of informal collaboration is generated from 854 papers published in either 1997, 1998 or 1999 and consists of 3,142 researchers. In comparison, the 2011 network connects 7,028 researchers that have collaborated on 1,889 papers. This is not a trivial result. Rather, it could have been the case that authors tend to collaborate informally with researchers already in the network, i.e. those other authors have previously informally collaborated with.

Table 2.4: Global network measures for the networks of informal collaboration.

	Overall				Giant				
	Size	Links	Avg. clustering	Components	Size	Density	Avg. path length	Diameter	rho
1999	3142	10639	0.099	30	3001	0.0023	4.59	12	0.48***
2000	3286	11171	0.099	33	3115	0.0022	4.62	13	0.49***
2001	3399	11382	0.112	34	3234	0.0021	4.67	13	0.49***
2002	3559	12071	0.105	37	3381	0.002	4.74	14	0.50***
2003	3815	13498	0.105	31	3671	0.0019	4.73	13	0.48***
2004	4191	15234	0.103	34	3989	0.0018	4.64	13	0.52***
2005	4521	16909	0.103	31	4388	0.0017	4.67	13	0.50***
2006	4838	17868	0.086	37	4693	0.0016	4.76	14	0.52***
2007	5265	20930	0.091	40	5102	0.0016	4.72	16	0.51***
2008	5725	23662	0.1	43	5535	0.0015	4.66	14	0.54***
2009	6220	28353	0.105	44	6016	0.0015	4.53	14	0.54***
2010	6661	30567	0.097	48	6445	0.0014	4.57	13	0.53***
2011	7028	33248	0.103	54	6783	0.0014	4.50	15	0.54***

Notes: Table presents global network statistics for all three-year co-author networks of informal collaboration. Each network connects researcher that have that collaborated formally (co-authoring) or informally on papers published in year t, $t-1$ or $t-2$. *Size* is the number of researchers in the network resp. largest component. *Links* is the number of links connecting the researchers. *Components* is the number of distinct network components. *Density* is the share of realized to potential paths (equation (2.7)) in the largest component. *Avg. path length* is the average length of all possible paths between any two researchers in the largest component. *Diameter* is the longest of all shortest paths between all researchers in the largest component. *Avg. clustering* is the average clustering coefficient of all nodes in the network's largest component (2.10). *rho* is the Spearman rank correlation coefficient between all researchers' betweenness centrality (2.5) and Eigenvector centrality (2.4) in the largest component, with ***, ** and * indicating statistical significance to the 10, 5 and 1 percent level.

Table 2.5: Global network measures for the co-author networks.

	Overall				Giant				
	Size	Links	Avg. clustering	Components	Size	Density	Avg. path length	Diameter	rho
1999	1201	966	0.374	481	33	0.0909	3.64	8	0.37**
2000	1255	1017	0.376	498	53	0.0581	5.16	13	0.30**
2001	1324	1082	0.374	524	45	0.0616	5.97	15	0.28*
2002	1416	1156	0.365	556	61	0.0443	6.26	15	0.35***
2003	1478	1237	0.371	565	66	0.0434	5.71	13	0.29**
2004	1659	1431	0.385	603	68	0.0435	6.14	14	0.35***
2005	1794	1581	0.409	647	131	0.0222	8.34	20	0.28***
2006	2044	1789	0.41	744	65	0.0529	3.71	8	0.53***
2007	2272	2171	0.455	747	264	0.0123	9.91	26	0.15**
2008	2551	2546	0.479	792	128	0.025	5.76	12	0.32***
2009	2762	2904	0.495	788	591	0.0054	12.26	34	0.15***
2010	2959	3063	0.497	860	505	0.0064	12.64	32	0.13***
2011	3109	3236	0.488	898	601	0.0053	10.76	27	0.20***

Notes: Table presents global network statistics for all three-year co-author networks. Each network connects researchers that have jointly published a paper in year t, $t-1$ or $t-2$. *Size* is the number of researchers in the network resp. largest component. *Links* is the number of links connecting the researchers. *Components* is the number of distinct network components. *Density* is the share of realized to potential paths (equation (2.7)) in the largest component. *Avg. path length* is the average length of all possible paths between any two researchers in the largest component. *Diameter* is the longest of all shortest paths between all researchers in the largest component. *Avg. clustering* is the average clustering coefficient of all nodes in the network's largest component (2.10). *rho* is the Spearman rank correlation coefficient between all researchers' betweenness centrality (2.5) and Eigenvector centrality (2.4) in the largest component, with ***, ** and * indicating statistical significance to the 10, 5 and 1 percent level.

Second, the co-author networks are sparse and unconnected in all years, but the inclusion of links of informal collaboration drastically improves connectivity: (i) There are up to 898 distinct components in the co-author networks but at most 54 in the networks of informal collaboration; (ii) In the co-author networks less than a fifth of all researchers are connected within one component; In the networks of informal collaboration, however, the largest component consistently captures at least 95% of all researchers; (iii) Average path length and diameters are usually lower in the largest component of the network of informal collaboration compared to the network of formal collaboration's largest component. Figure 2.1 visualizes the dramatic increase in connectedness due to the inclusion of links representing informal collaboration. From left to right, i.e. going from the 1997-1999 period to the 2009-2011 period, the network size and number of components increase. Going from top to bottom, i.e. going from co-author networks to networks of informal collaboration, the number of nodes increases, the number of components decreases, and the size of the giant component increases. These differences in network size and connectivity have implications for the study of the flow of information or spillovers due to collaboration.

Third, despite becoming more inclusive, both networks become less dense over time. Looking at the network of informal collaboration only, density decreased from 0.0023 in 1999 to 0.0014 in 2011. This is because the growth rate of links (in this case a proxy for collaborations) does not keep up with the growth rate of the number of participating researchers.

Fourth, a direct comparison suggests that for our period of analysis, the social network of informal collaboration always exhibits small-world properties, while the social network of formal collaboration never. Small-world networks, whose name is based on the small-world phenomenon, have unique information transfer capabilities (Watts and Strogatz, 1998; Watts, 1999). A network with small world properties is one with a high clustering, a small average distance, a high number of nodes as compared to the number of links, and the fact that a large share of authors are somehow connected via a series intermediate steps (i.e. a giant component exists). Goyal et al. (2006) report how the world of academic economists has only become a small world by the 1990-2000 period, as compared to the two previous decades.[22] We find that

[22]The authors study co-author networks covering 10 years of publications in up to 105 journals listed in EconLit.

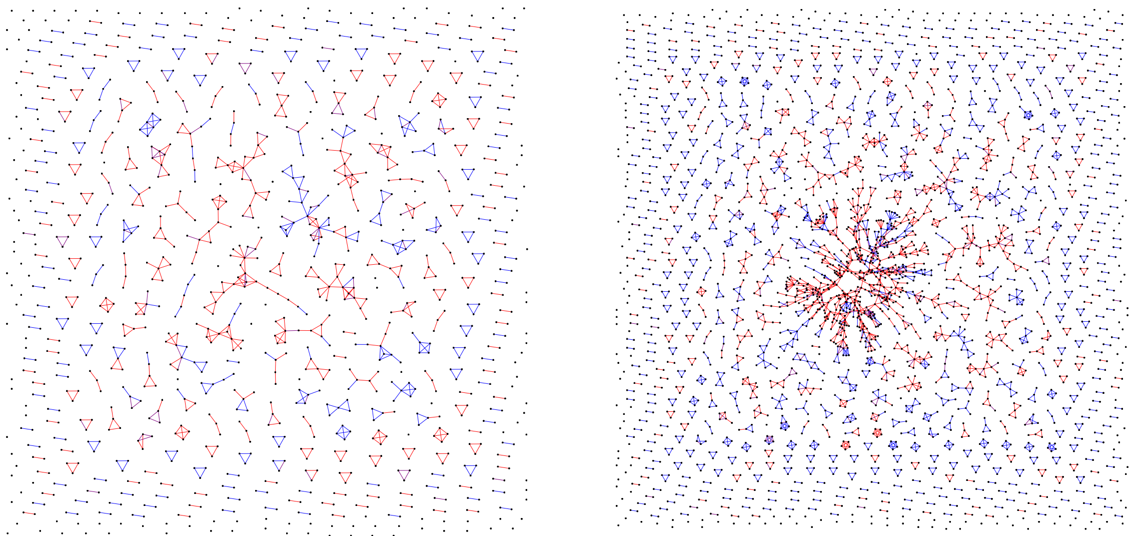
none of the networks under consideration in our study display small-world properties: The co-author networks have high average distance and giant component does not exist, while for the networks of informal collaboration clustering is too low.

Fifth, in the network of informal collaboration, betweenness centrality and eigenvector centrality of the same researcher do not strongly correlate. This is indicated by the Spearman correlation coefficient ("rho") between eigenvector centrality and betweenness centrality in the last column of table 2.4, which never exceeds 0.54. Researchers that are important for the flow of information (high betweenness centrality) are not often also well-suited to influence the network (high eigenvector centrality).

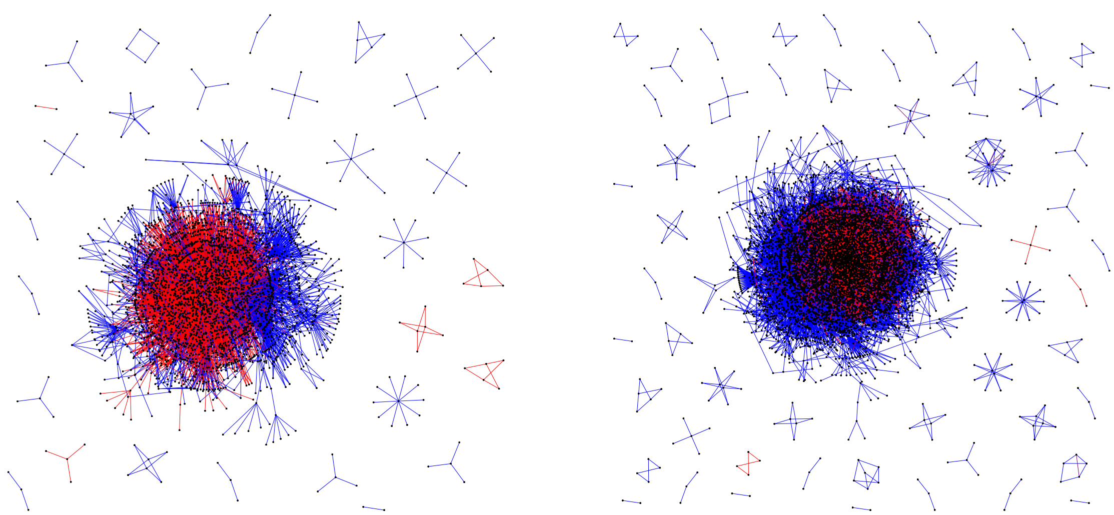
Sixth, figure 2.1 furthermore shows a hierarchy in the flow of information. Links are colored according to which journal an article was published in: Red indicates a top journal publication, while blue links indicate that the paper was published in one of the three other journals. The few links occurring in both groups of journals are colored purple. For both networks it holds that the center of the network is dominated by links that originate from top journals. This is indicative of a higher connectedness of the collaborators involved in publications in top journals.

Figure 2.1: Comparison of networks of informal and formal collaboration, 1997-1999 and 2009-2011.

(a) Network of formal collaboration, 1997-1999 (b) Network of formal collaboration, 2009-2011



(c) Network of informal collaboration 1997-1999 (d) Network of informal collaboration, 2009-2011

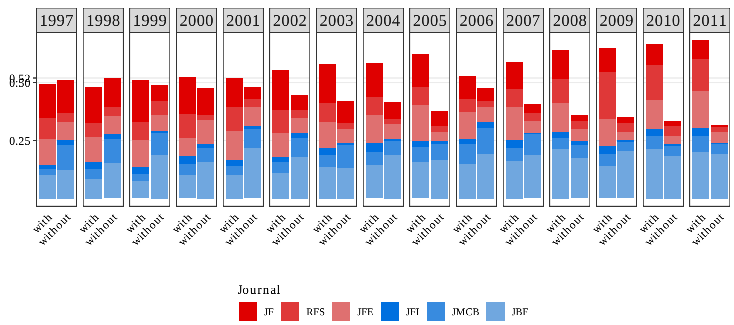


Notes: Figures show networks of formal collaboration (top) and network of informal collaboration (bottom) for publications in six financial economics journals published between 1997 and 1999 (left) and 2009 and 2011 (right). In the network of formal collaboration, a link is drawn between every author of a published paper. In the network of informal collaboration, a link is drawn between an acknowledged commenter and every author of a published paper. Red links indicate that the paper was published in a top journal, while blue indicates a publication on other journals. If a link occurs in both a top journal and other journals, which is a rare event, the link is colored in purple. Position in the network was computed according to the Fruchterman-Reingold algorithm.

2.3.4 Intensive and Extensive Margin of Informal Collaboration

The vast majority of published research papers acknowledges informal input by commenters. Of the 6,401 papers in our sample, 5,641 (≈ 90%) papers acknowledge at least one commenter, one seminar, or one conference (Figure 2.2).[23] While we do not find much time-variation in this share, there is a substantial variation on the journal-level (figure 2.2). In top journals almost every paper reports at least one form of informal collaboration.

Figure 2.2: Share of papers with and without acknowledgements, by journal and year.

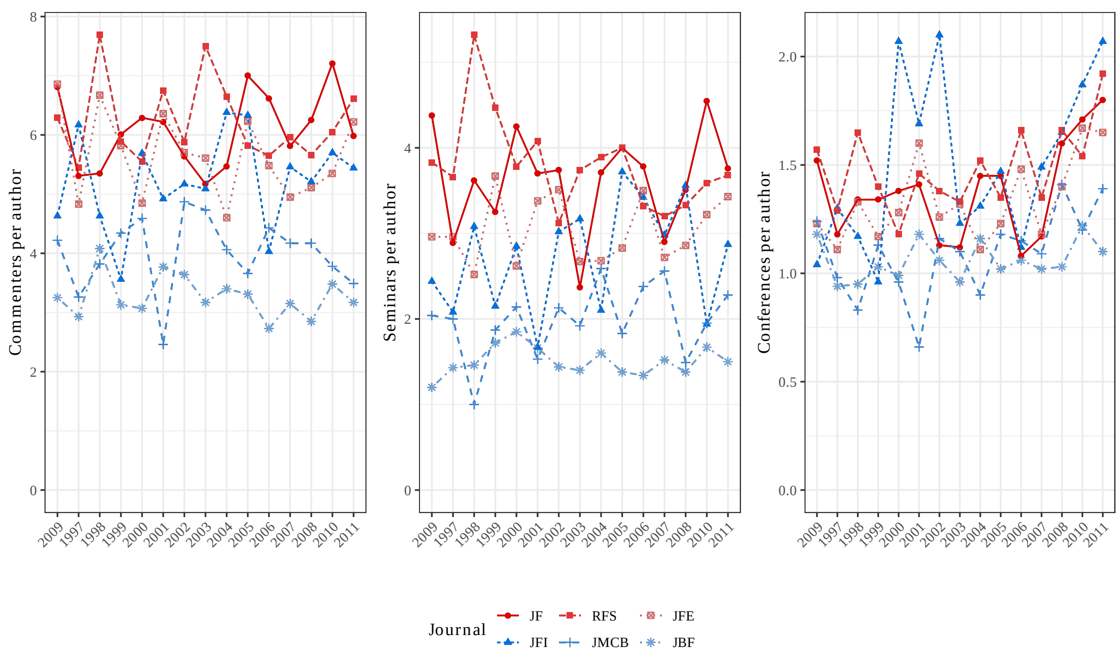


Notes: Graph shows share of papers with (left bar) and without acknowledgment section (right bar) for each year. An acknowledgement may contain named researchers mentioned for feedback, advice and discussion (unless she's the journal's managing editor), universities or conferences where the paper was presented. Colors correspond to journals, where red-ish colors refer to the three top journals (JF, JFE, RFS) and blue-ish colors refer to the three other journals (JFI, JMCB, JBF).

Top journal publications not only acknowledge informal collaboration more often, they also report a higher intensity thereof. On average, top journal publications acknowledge almost twice as many commenters as papers published by the other journals, and are presented more than twice as often at seminars and conferences (Figure 2.3). It is remarkable how well the ranking of these six journals according to their impact factors is reflected in the average intensity of informal collaboration. Another interesting feature is the similarity of the JFI (blue dotted line) to the group of top journals (red lines) in terms of informal collaboration. One plausible explanation is that papers, that aimed for the top journals but got rejected were then submitted to the JFI.

[23] The remaining papers may acknowledge the editor, anonymous referees, funding, data exchange and research assistance.

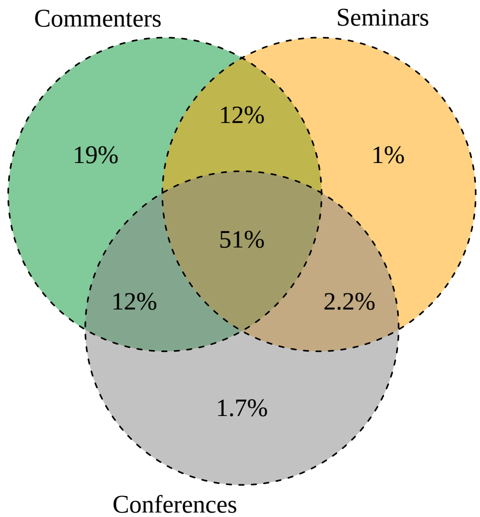
Figure 2.3: Mean number of author-normalized commenters, seminars and conferences over time per journal and year.



Notes: Graph shows mean number of acknowledged commenters (without the journal's managing editor) (left plot), seminars (center plot) and conferences (right plot) per journal over time, divided by the number of authors. Colors correspond to journals, where red-ish colors refer to the three top journals (JF, JFE, RFS) and blue-ish colors refer to the three other journals (JFI, JMCB, JBF).

Finally the three types of informal collaboration do not appear to be substitutes to each other, as almost half of all the papers report all three forms, as the Venn diagram in figure 2.4 shows, and very few papers report only one form.

Figure 2.4: Share of papers jointly acknowledging researchers by name, seminars and conferences.



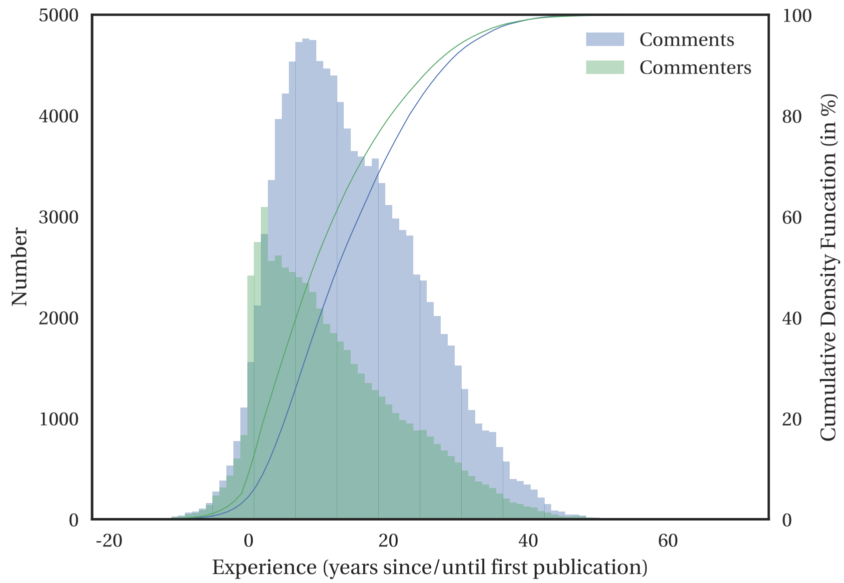
Notes: Venn diagram shows share of papers that acknowledge commenters (excluding the journal's managing editor), seminars and conferences.

2.3.5 Life-cycle effects and Reciprocity

Another interesting question about informal collaboration in financial economics is to ask who is being acknowledged. Giving comments to authors is not evenly distributed across a researcher's life-cycle. The two histograms in figure 2.5 and the accompanying cumulative density functions make it clear that the majority of comments are given by researchers with 3 to 20 years of academic experience. This finding is irrespective of whether we use the simple count of commenters at a given academic experience (green bars and green line) or whether we weigh the count with the number of given comments in a year (blue bars and blue line). If we do not weigh the commenters by the number of commenters, we find the mode commenter to have 2 years of academic experience. In the weighted case, the modal commenter has 7 years of academic experience.

Figure 2.6 tracks for each commenter the number of papers acknowledging her over her academic lifecyle. Very few researchers are acknowledged at a very high age. Among them are Milton Friedman and Paul Samuelson, who are represented by the rightmost lines.

Figure 2.5: Histogram and CDF for number of commenters and comments by academic experience.

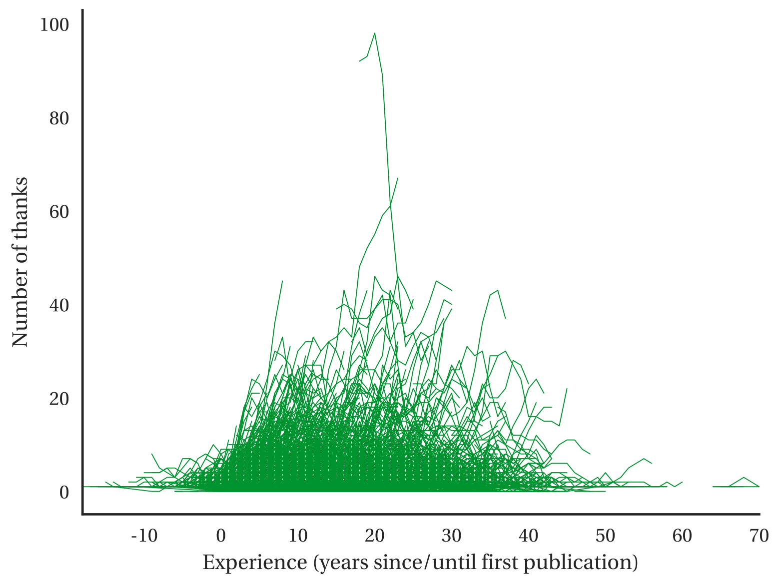


Notes: Histogram showing the number of commenters (green) and comments (blue) by academic experience on the left axis. Right axis shows corresponding cumulative distribution functions in percent. Experience is the number of years between first publication, as measured by Scopus, and the publication year of the paper that acknowledges the commenter.

Figure 2.7 plots the joint density of author experience and commenter experience, measured in the year of the publication of the paper. The dashed line indicates a line of equal age: If authors would mainly acknowledge commenters of the same age we would expect the mass of the joint distribution along this line. This is not the case: The highest mass is for authors with no experience and commenters with between three and 11 years since their first publication. These are authors who publish for the first time. The joint distribution has a fatter upper tail along the commenter's experience as opposed to the author's experience. The reason is that it is more common to find informal collaboration where the author is less senior than the commenter rather than the other way around.

The observation that commenting on each others work is prevalent raises the question as to why researchers invest their scarce time to read manuscripts when they do not receive tangible credit for it. One possible explanation is that researchers are following a quid-pro-quo

Figure 2.6: Number of comments in our dataset over the academic lifecycle, by researcher.

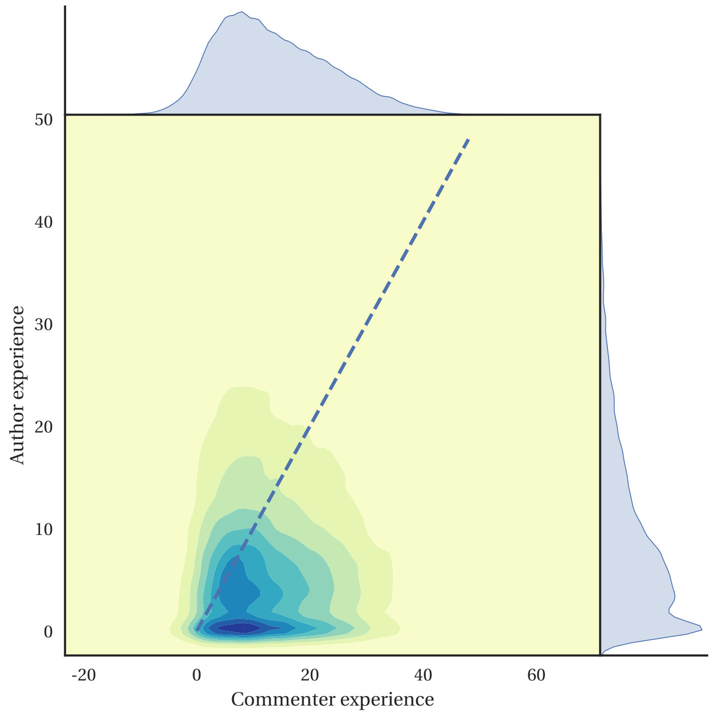


Notes: Plot shows the relation between the number of papers that acknowledge a given researcher to her experience. Experience is the number of years between first publication, as measured by Scopus, and the publication year of the paper that acknowledges the commenter.

strategy, that is to help researchers who have helped them. Reciprocity in informal collaboration takes two forms: First, comment on the work of one's coauthors, and second, comment on the work of one's commenters. Accordingly, we define a paper a as exhibiting reciprocity if it satisfies one of two conditions: 1) An acknowledged commenter is a co-author of at least one of the authors of paper a; 2) An acknowledged commenter publishes a paper in our dataset not jointly with an author in κ_a, and one of the authors of a is acknowledged. Of course, for condition 1) or 2) to hold, at least one of the acknowledged commenters must be an author in the dataset, and at least one paper is not single-authored. For this reason we only consider papers with this prerequisite.

We find 455 papers that acknowledge informal input from co-authors of other papers—out of 5,031 papers where this is possible, i.e. where at least one of the author's coauthors is an author in our dataset. There are 3,098 papers where at least one acknowledged commenter acknowledges at least one of the authors on her papers. To put this into perspective, there are

Figure 2.7: Heatmap showing joint distribution of author's experience and commenter's experience.



Notes: Heatmaps displays the joint distribution of academic experience of authors (left axis, right marginal plot) and of academic experience of commenters (bottom axis, top marginal plot). Dark areas indicate higher density. Experience is the number of years between first publication, as measured by Scopus, and the publication year of the paper that acknowledges the commenter.

5,012 papers where we could observe reciprocity by acknowledged commenters, i.e. where at least one commenter authors a paper without the authors who acknowledged her. In total, 3,214 (about 70% of 5,031) papers fulfill one of the above reciprocity conditions.

One caveat is in order, as we only observe commenter links within a set of six Financial Economics journals. This is not necessarily the natural domain of all commenters. For example, the Nobel laureate in Economics in 2014, Lars Peter Hansen, has been acknowledged by more than 20 papers in our dataset, while he didn't author one paper in our dataset. The same holds true for the 1993 Nobel laureate Douglass North, who appears in two acknowledgement sections. Both might well acknowledge any of the authors that acknowledged them in papers published in other journals.

2.3.6 Covariates of Centrality Rank and Number of Acknowledgments

Next, we want to understand how centralities in the networks are related to one another and–more importantly–with observable characteristics of researchers. Considering researcher i in the network corresponding to year t, we thus estimate the following empirical model

$$\begin{aligned}\text{Centrality}_{i,t} = &\beta_0 + \beta_1\text{female}_i + \beta_2\text{Euclid}_{it} + \beta_3\text{PublicationStock}_{it} + \\ &\beta_3\text{CitationStock}_{it} + \beta_4\text{Experience}_{it} + \beta_4\text{ExperienceSQ}_{it} + \mathbf{D}_t + \epsilon_i \end{aligned} \tag{2.12}$$

where *Centrality* is one of eigenvector centrality rank, betweenness centrality rank, out-degree, and number of thanks. Though correlated, out-degree and the number of thanks differ: out-degree measures the number of authors acknowledging a researcher, number of thanks the number of papers. This accounts for the fact that the commenter possibly spoke to multiple authors of a paper on which she is acknowledged. Each variable is measured in the network corresponding to year t, i.e. the network inferred from acknowledged commenters on papers published in t, $t-1$, and $t-2$. All other variables are measured in t. We include fixed effects for t to account for the growth of the network over time. We cluster standard errors on the researcher level to capture unobserved heterogeneity. There are two sources of unobserved heterogeneity. One is a different frequency in the data: Some researcher occur in all networks, while others only in one network. The other source of unobserved heterogeneity is different individual networks that authors can draw from. If the dependent variable is a rank (for eigenvector centrality rank and betweenness centrality rank), a negative β indicates a positive relationship.

Table 2.6 presents summary statistics for the sample. Since we pool all researchers in all networks for all years, there are 65,800 observations (where the same researcher can show up with different centrality in different years). Table 2.7 presents Spearman correlation coefficients of the variables. It is noteworthy how weakly eigenvector centrality rank correlates with author characteristics, as no coefficient surpasses -0.12.[24] The number of thanks correlates with productivity measures only weakly, too: Of all the Spearman correlations between number

[24]A negative correlation indicates a positive relationship between better centrality ranks (lower numbers) and productivity.

of thanks and any of the author metrics, the highest is with the Euclidean index of citations and equals 0.37. Being central in the network of informal collaboration is hence not the same as being a prolific author.[25]

Table 2.6: Summary statistics for all variables continued variables used in the pooled centrality sample.

	N	Mean	Median	Std.Dev.	Min	Max
Characteristics						
Euclid. Index	56997	98.8	27	265.38	0	9676
Publication stock	56997	13.8	8	17.75	0	339
Citation stock	56997	1712.4	180	5961.26	0	326129
Experience	56997	11.8	10	10.38	−18	70
Experience SQ	56997	247.8	100	364.82	0	4900
Informal Collaboration						
Eigenvector centrality	51893	2358.3	2139	1612.04	1	6783
Betweenness centrality	51893	2232.5	2070	1492.46	1	5766
Out-Degree	56997	4.0	2	6.64	0	123
No. of Thanks	56997	2.0	1	3.40	0	98

Notes: Summary statistics for pooled centrality sample, where the unit of observation is the combination of researcher i and year t. *Euclid. Index* is the Euclidean index of citations of a researcher (equation (2.1)) in year t. *Publication Stock* is the count of all publications published until year t (including). *Citation Stock* is the count of all citations to all publications until year t (including). *Experience* is the number of years between the year of first publication and t. *Experience SQ* is its square. These five variables are computed using data from Scopus. *Eigenvector centrality rank* is the rank according to the researcher's Eigenvector centrality in the network corresponding to year t. *Betweenness centrality rank* is the rank according to the researcher's betweenness centrality in the network corresponding to year t. *Out-Degree* is the number of distinct authors that acknowledge this researcher in t, $t-1$ and $t-2$. *No. of Thanks* is the number of papers published in t, $t-1$ and $t-2$ that acknowledge this researcher.

Table 2.8 presents results of a pooled OLS regression for model (2.12) with different dependent variables. The first striking observation is that females are acknowledged less often and have a lower out-degree, even at the same level of academic productivity and experience. Female researchers also rank worse in terms of eigenvector centrality ($\approx$ 91 ranks) and betweenness centrality ($\approx$ 120 ranks), but only when excluding the year-fixed effects in models (2) and (4). The statistical malus for female researchers is hence time-varying. The Euclidean index of citations is statistically significantly associated with higher betweenness centrality ranks, higher

[25] These correlations hold for the pooled sample. Looking at period-wise levels, we find the same pattern.

Table 2.7: Correlation coefficients for all continued variables used in the regression as well as co-author network variables.

Characteristics								
Euclid. Index								
Publication stock	0.81							
Citation stock	0.97	0.87						
Experience	0.77	0.81	0.87					
Experience SQ	0.75	0.78	0.86	0.98				
Informal Collaboration								
Eigenvector centrality	0.02	0.05	0.06	0.14	0.16			
Betweenness centrality	-0.10	-0.10	-0.07	0.02	0.04	0.64		
Out-Degree	0.36	0.21	0.33	0.23	0.25	-0.02	-0.31	
No. of Thanks	0.38	0.23	0.35	0.25	0.26	-0.05	-0.33	0.93

Notes: Spearman correlation coefficients for pooled centrality sample, where the unit of observation is the combination of researcher i and year t. *Euclid. Index* is the Euclidean index of citations of a researcher (equation (2.1)) in year t. *Publication Stock* is the count of all publications published until year t (including). *Citation Stock* is the count of all citations to all publications until year t (including). *Experience* is the number of years between the year of first publication and t. *Experience SQ* is its square. These five variables are computed using data from Scopus. *Eigenvector centrality rank* is the rank according to the researcher's Eigenvector centrality in the network corresponding to year t. *Betweenness centrality rank* is the rank according to the researcher's betweenness centrality in the network corresponding to year t. *Out-Degree* is the number of distinct authors that acknowledge this researcher in t, $t-1$ and $t-2$. *No. of Thanks* is the number of papers published in t, $t-1$ and $t-2$ that acknowledge this researcher.

out-degree and a higher number of thanks. Publication stock is statistically significantly associated with higher eigenvector centrality ranks and higher betweenness centrality ranks. Citation stock is only statistically significantly associated with lower betweenness centrality ranks. There are interesting life-cycle effects at work. More experienced authors tend to be less eigenvector central (one more year $\approx$ -16 eigenvector centrality ranks), have a higher out-degree and are thanked more often, but they are not more betweenness central. The relationship between experience and eigenvector centrality rank appears to be increasing, the relationships with the other variables are decreasing. Finally, as expected, a higher number of thanks is statistically significantly associated with higher centrality ranks.

Table 2.8: Regression results for pooled OLS estimation explaining centrality in the network of informal collaboration and number of thanks.

	Eigenvector centrality rank		Betweenness centrality rank		Out-Degree		No. of Thanks	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female	-17.769	91.218^{***}	8.835	120.675^{***}	-0.542^{***}	-0.499^{***}	-0.295^{***}	-0.299^{***}
	(29.799)	(31.513)	(28.130)	(30.201)	(0.171)	(0.169)	(0.084)	(0.084)
Euclid. Index	-0.044	0.265^{**}	-0.194^{**}	0.124	0.006^{***}	0.007^{***}	0.003^{**}	0.003^{**}
	(0.088)	(0.106)	(0.087)	(0.123)	(0.002)	(0.002)	(0.001)	(0.001)
Publication stock	-1.723^{**}	0.434	-6.975^{***}	-4.759^{***}	0.002	0.003	0.003	0.003
	(0.756)	(0.852)	(0.819)	(0.900)	(0.010)	(0.010)	(0.005)	(0.005)
Citation stock	0.002	-0.005	0.012^{***}	0.005	0.00002	0.00001	0.00002	0.00002
	(0.004)	(0.005)	(0.004)	(0.005)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Experience	16.028^{***}	17.564^{***}	-1.179	0.379	0.202^{***}	0.202^{***}	0.101^{***}	0.101^{***}
	(2.384)	(2.789)	(2.537)	(2.977)	(0.022)	(0.022)	(0.011)	(0.011)
Experience SQ	0.235^{***}	0.226^{***}	0.547^{***}	0.539^{***}	-0.005^{***}	-0.005^{***}	-0.003^{***}	-0.003^{***}
	(0.065)	(0.078)	(0.071)	(0.085)	(0.001)	(0.001)	(0.0004)	(0.0004)
No. of Thanks	-96.683^{***}	-96.934^{***}	-157.451^{***}	-157.698^{***}				
	(6.424)	(5.776)	(12.489)	(11.710)				
Constant		$2,261.929^{***}$		$2,458.945^{***}$		2.325^{***}		1.148^{***}
		(21.457)		(25.287)		(0.082)		(0.042)
Year-FE	*Yes*	*No*	*Yes*	*No*	*Yes*	*No*	*Yes*	*No*
Clustered SE	Individual	Individual	Individual	Individual	Individual	Individual	Individual	Individual
Mean	2358	2358	2232	2232	4.01	4.01	2	2
N	51,893	51,893	51,893	51,893	56,997	56,997	56,997	56,997
Adjusted R^2	0.199	0.062	0.311	0.144	0.099	0.098	0.097	0.097

Notes: Standard errors in parenthesis. ***, ** and * indicate statistical significance to the 1, 5 and 10 percent level, respectively. *Eigenvector centrality rank* is the rank according to the researcher's Eigenvector centrality in the network of informal collaboration corresponding to year t (equation (2.4)). *Betweenness centrality rank* is the rank according to the researcher's betweenness centrality in the network of informal collaboration corresponding to year t (equation (2.5)). *Out-Degree* is the number of distinct authors that acknowledge this researcher in t, $t-1$ and $t-2$. *No. of Thanks* is the number of papers published in t, $t-1$ and $t-2$ that acknowledge this researcher. *Female* indicates that the researcher's first name is estimated to be female. *Euclid. Index* is the Euclidean index of citations in year t. *Publication Stock* is the count of all publications published until year t (including). *Citation Stock* is the count of all citations to all publications until year t (including). *Experience* is the number of years between the year of first publication and t. *Experience SQ* is its square. These five variables are computed using data from Scopus. These five variables are computed using data from Scopus.

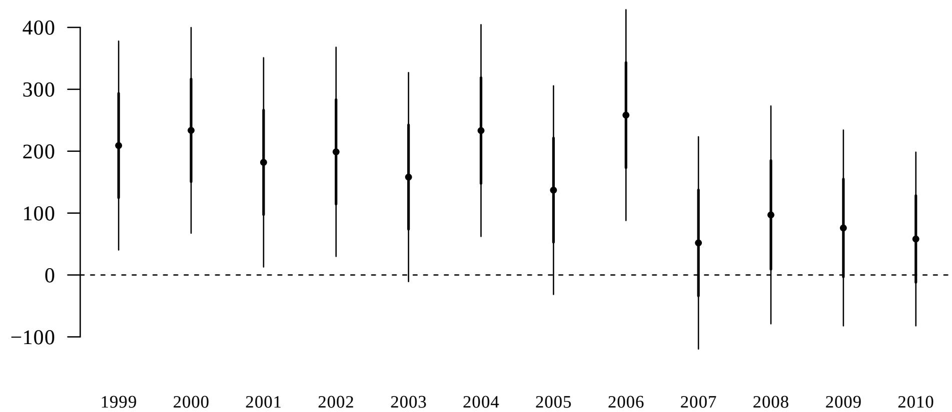
Table 2.9 shows that the malus for female academics is vanishing over time in our dataset– female researchers are no longer significantly less central than their male counterparts. The table shows the coefficients for interactions of the female variable and the year variable, relative to year 2011. In both models, we include all individual characteristics and cluster standard errors around the indivdual researcher. After 2006, the interaction coefficient is not statistically significant anymore. Put differently, past 2006 we don't find evidence of a female malus for eigenvector centrality rank in the network of informal collaboration. For betweenness centrality the last year with signs of a statistical malus is 2005, although the malus has always been weak. Figure 2.8 plots the time-varying relationship and thus gives a better grasp of the magnitude.

For comparison we conduct a similar regression for eigenvector centrality rank, betweenness centrality rank and degree measured in the network of formal collaboration. The results are reported in table 2.10. The number of observations for the first two columns (eigenvector centrality rank and betweenness centrality rank) is very low because we only consider centrality ranks of authors in the networks' giant components. The first row indicates that female authors do not rank worse. The Euclidean index of citations impacts eigenvector centrality rank and betweenness centrality rank negatively. Authors that publish more are more eigenvector- and betweenness central, and also have a higher degree.[26] We find that citation stock furthermore negatively affects degree, i.e. authors who are cited more have fewer distinct co-authors holding constant the number of papers they have published. More experienced authors also have more distinct coauthors, albeit at a decreasing rate. This can be seen from the positive coefficient for Experience and the negative coefficient for Experience^2. Finally, authors who are more often thanked also tend to be more eigenvector central, more betweenness central and have a higher degree: For each additional paper that thanks the author, the author is expected to be 1 rank position more eigenvector central, 3 rank positions more betweenness central, and have 0.1 more distinct co-authors.

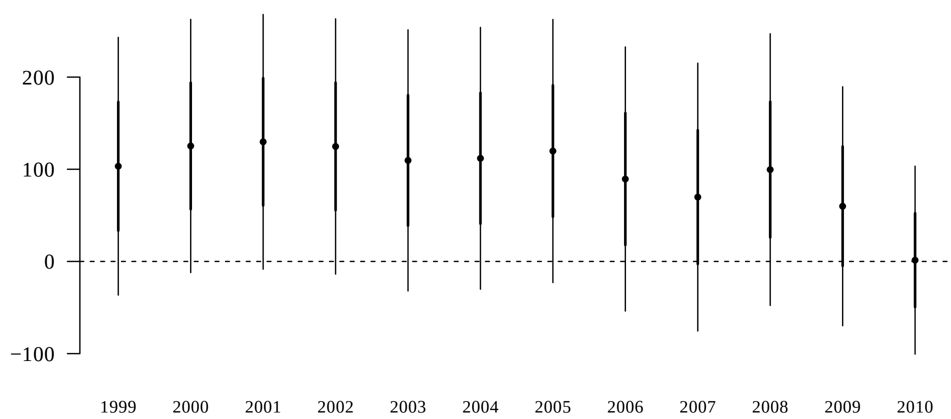
[26]These effects may partly be mechanical since publishing often is correlated with having distinct coauthors.

Figure 2.8: Coefficient plot for different centrality values of female researchers interacted with years.

(a) Eigenvector centrality rank



(b) Betweenness centrality rank



Notes: Figures depict coefficients and standard deviations from table 2.9. Coefficients are interactions of female*year, with 2011 as reference category. *Eigenvector cent. rank* is the researcher's eigenvector centrality in the giant component of the network of informal collaboration according to equation (2.4). *Betweenness cent. rank* is the researcher's betweenness centrality in the giant component of the network of informal collaboration according to equation (2.5).

Table 2.9: Interaction effects of Female*year of pooled OLS estimation explaining centrality in the network of informal collaboration.

	Eigenvector centrality rank	Betweenness centrality rank
	(1)	(2)
Female	-138.444^{**}	-64.957
	(67.295)	(57.513)
1999:Female	209.111^{**}	103.306
	(84.380)	(69.931)
2000:Female	233.652^{***}	125.274^{*}
	(83.075)	(68.704)
2001:Female	181.974^{**}	129.789^{*}
	(84.542)	(69.110)
2002:Female	198.958^{**}	124.757^{*}
	(84.514)	(69.257)
2003:Female	158.170^{*}	109.637
	(84.446)	(70.848)
2004:Female	233.346^{***}	111.904
	(85.576)	(71.031)
2005:Female	137.089	119.801^{*}
	(84.299)	(71.375)
2006:Female	258.295^{***}	89.434
	(85.162)	(71.662)
2007:Female	51.851	69.831
	(85.747)	(72.676)
2008:Female	97.067	99.653
	(88.003)	(73.712)
2009:Female	75.983	59.915
	(79.120)	(64.836)
2010:Female	58.139	1.386
	(70.158)	(50.995)
Controls	*Yes*	*Yes*
Clustered SE	Individual	Individual
Mean	2358	2232
N	51,893	51,893
Adjusted R^2	0.199	0.311

Notes: Standard errors in parenthesis. ***, ** and * indicate statistical significance to the 1, 5 and 10 percent level, respectively. *Eigenvector centrality rank* is the rank according to the researcher's Eigenvector centrality in the network of informal collaboration corresponding to year t (equation (2.4)). *Betweenness centrality rank* is the rank according to the researcher's betweenness centrality in the network of informal collaboration corresponding to year t (equation (2.5)). *Female* indicates that the researcher's first name is estimated to be female.

Table 2.10: Regression results for pooled OLS estimation explaining centrality in the network of formal collaboration.

	Eigenvector centrality rank		Betweenness centrality rank		Degree	
	(1)	(2)	(3)	(4)	(5)	(6)
Female	−6.045	9.092	−1.308	8.620	0.014	0.048
	(9.172)	(11.793)	(3.823)	(6.058)	(0.039)	(0.039)
Euclid. Index	0.001	0.063*	−0.007	0.034*	0.0002	0.0004
	(0.024)	(0.036)	(0.012)	(0.018)	(0.0003)	(0.0003)
Publication stock	−0.802***	−1.222***	−0.886***	−1.183***	0.013***	0.014***
	(0.304)	(0.396)	(0.179)	(0.223)	(0.002)	(0.002)
Citation stock	0.001	0.001	0.001**	0.001*	−0.00002*	−0.00003*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.00001)	(0.00001)
Experience	0.091	−0.088	−0.395	−0.487	0.014**	0.011*
	(1.151)	(1.523)	(0.578)	(0.876)	(0.006)	(0.006)
Experience SQ	−0.006	−0.001	0.036*	0.038	−0.001***	−0.0005**
	(0.032)	(0.042)	(0.019)	(0.026)	(0.0002)	(0.0002)
No. of Thanks	−0.538	−1.066*	−2.383***	−2.712***	0.087***	0.086***
	(0.446)	(0.622)	(0.598)	(0.428)	(0.006)	(0.006)
Constant		220.080***		159.506***		1.463***
		(9.527)		(5.281)		(0.026)
Year-FE	*Yes*	*No*	*Yes*	*No*	*Yes*	*No*
Clustered SE	Individual	Individual	Individual	Individual	Individual	Individual
Mean	210	210	140	140	1.87	1.87
N	2,611	2,611	2,611	2,611	25,824	25,824
Adjusted R^2	0.385	0.011	0.620	0.067	0.130	0.116

Notes: Standard errors in parenthesis. ***, ** and * indicate statistical significance to the 1, 5 and 10 percent level, respectively. *Eigenvector centrality rank* is the rank according to the researcher's Eigenvector centrality in the network of formal collaboration corresponding to year t (equation (2.4)). *Betweenness centrality rank* is the rank according to the researcher's betweenness centrality in the network of formal collaboration corresponding to year t (equation (2.5)). *Degree* is the researcher's degree in the network of formal collaboration corresponding to year t (equation (2.2)). *Female* indicates that the researcher's first name is estimated to be female. *Euclid. Index* is the Euclidean index of citations in year t. *Publication Stock* is the count of all publications published until year t (including). These five variables are computed using data from Scopus. *Citation Stock* is the count of all citations to all publications until year t (including). *Experience* is the number of years between the year of first publication and t. *Experience SQ* is its square.

Centrality in the network of informal collaboration is only mildly correlated with existing measures of academic influence such as publication count or a joint measure of citations and publications. We therefore argue that centralities provide an alternative ranking method to assess the influence a scholar has on the profession.

2.3.7 Who are the most central authors and commenters?

Laband and Tollison (2003) compile a list of the most often thanked authors from a sample of three general interest Economics journals over forty years. Our sample however uses a more homogeneous sample[27] and adopts a network view. Table 2.11 ranks researchers based on their average rankings according to different measures (the full list is available online). We rank researchers according to how often they have been acknowledged, their betweenness centrality in the co-author network, their betweenness centrality in the commenter network, eigenvector centrality in the co-author network and their eigenvector centrality in the commenter network. We add corresponding statistics derived from the co-author network as a contrast.

Some of the greatest financial economists of our time are prominently featured in the ranking. Stulz, R. M. has been acknowledged most often, followed by Stein, J. C. and Ritter, J. R. Both Stein, J. C. and Stulz, R. M. are also very eigenvector central: Stulz, R. M. is the most eigenvector central in the Co-Author network (followed by Berger, A. N. and Titman, S. D.), while Stein, J. C. is the most eigenvector central researcher in the Commenter network (followed by Shleifer, A. and Zingales, L.). However, other researchers are most betweenness central: In the Co-Author network, Shivdasani, A., Chen, J. and Lemmon, M. L. are the most betweenness central researchers. In the Commenter network, the three most betweenness central researchers are Lin, C., Liu, J. and Ma, Y.

An interesting observation is the high placements of editors. Note that we do not count informal collaboration with editors on papers that fall within her tenure; all counts of acknowledgements of editors and the resulting links in the network of informal collaboration thus result

[27] General interest journals in Economics publish papers from a wider range of topics than do journals in Financial Economics.

Table 2.11: Top 30 researchers according to average rankings according to different centrality measures in all co-author and commenter networks.

	Thanks	Betweenness centrality		Eigenvector centrality	
		Co-Author network	Commenter network	Co-Author network	Commenter network
1	Stulz, R. M.	Shivdasani, A.	Lin, C.	Stulz, R. M.	Sensoy, B. A.
2	Stein, J. C.	Chen, J.	Liu, J.	Berger, A. N.	Yun, H.
3	Ritter, J. R.	Lemmon, M. L.	Ma, Y.	Titman, S. D.	Korteweg, A.
4	Shleifer, A.	Altman, E. I.	Cull, R.	Shleifer, A.	Stulz, R. M.
5	Titman, S. D.	Okunev, J.	Xuan, Y.	Ritter, J. R.	Kim, C.
6	Campbell, J. Y.	Cao, C. Q.	Cummins, J. D.	Harvey, C. R.	Hsu, P. H.
7	Amihud, Y.	Chordia, T.	Lim, T.	Flannery, M. J.	Xuan, Y.
8	Zingales, L.	Goldstein, I.	Weiss, M. A.	Graham, J. R.	Chen, H.
9	Green, R. C.	Liu, J.	Walter, I.	Ferson, W. E.	Ghent, A. C.
10	Ferson, W. E.	Cooney, J. W.	Lin, P.	Zingales, L.	Lyon, J. D.
11	Harvey, C. R.	Walter, I.	Clarke, G. R.	Karolyi, G. A.	Baker, M. P.
12	Duffie, J. D.	Haubrich, J. G.	Chen, J.	Amihud, Y.	Duchin, R. A.
13	Fama, E. F.	Hancock, D. G.	Zou, H.	Duffie, J. D.	Wurgler, J.
14	Jagannathan, R.	Ryngaert, M. D.	Scalise, J. M.	Stein, J. C.	Purnanandam, A. K.
15	Petersen, M. A.	Lo, A.	Hancock, D. G.	Subrahmanyam, A.	Sevick, M.
16	Schwert, G. W.	Brav, A.	Mithal, S.	Hirshleifer, D.	Kim, Y. C.
17	Flannery, M. J.	Subrahmanyam, A.	Zi, H.	Saunders, A.	Zhang, L.
18	Brennan, M. J.	Mester, L. J.	Neis, E.	Campbell, J. Y.	Tsai, C. L.
19	Rajan, R. G.	Stulz, R. M.	Kashyap, A. K.	Fohlin, C.	Chava, S.
20	Berger, A. N.	Kang, J. K.	Haubrich, J. G.	Boudreaux, D. J.	Seru, A.
21	French, K. R.	Davidson, I. R.	Song, F. M.	Khan, M. A.	Laeven, L.
22	Cochrane, J. H.	Berlin, M.	Wang, A. W.	Petersen, M. A.	Woo, S. J.
23	Daniel, K. D.	Sias, R. W.	Cao, C. Q.	Levine, R. L.	Subrahmanyam, A.
24	Allen, F.	Moon, C.	Covitz, D. M.	Brav, A.	Graham, J. R.
25	Kaplan, S. N.	Gosnell, T. F.	Piazzesi, M.	Ongena, S.	Roussanov, N.
26	Diamond, D. W.	Chiang, R. C.	Bonime, S. D.	Woo, D.	Tian, X.
27	Karolyi, G. A.	Hughes, J. P.	Ahn, H.	Lemmon, M. L.	Van Hemert, O.
28	O'Hara, M.	Berger, A. N.	Lo, A.	Servaes, H.	Kuehn, L. A.
29	Scharfstein, D. S.	Cull, R.	Michael, F. A.	Weisbach, M. S.	Knoeber, C. R.
30	Thakor, A. V.	Wilson, B. K.	Liu, P.	Starks, L. T.	Huang, J.

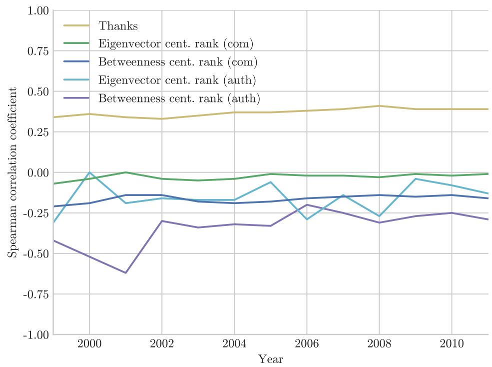
Notes: Table ranks researchers based on their average ranking according to various measures derived from publications in six financial economics journals published between 1997 and 2011. "Thanks" is the number of publications in this period that acknowledge the researcher for feedback, unless she was managing editor of the journal the paper got published in. "Betweenness centrality" (equation (2.5)) and "Eigenvector centrality" (equation (2.4)) are measured in the network connecting co-coauthors only resp. authors and commenters.

from papers published in journals other than those the editor is serving for. Editors are thus approached by other researchers even when the paper is not being published in their journal, suggesting an exposed role for editors in the profession (Brogaard et al., 2014).[28]

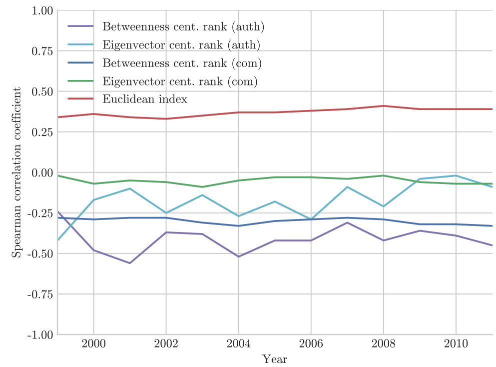
[28]An alternative explanation is that editors are being acknowledged even in the case of a desk-rejection on the paper.

Figure 2.9: Spearman rank correlation coefficients over period between various variables.

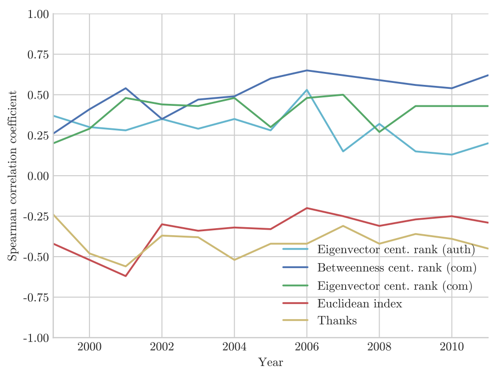
(a) Euclidean index of citations



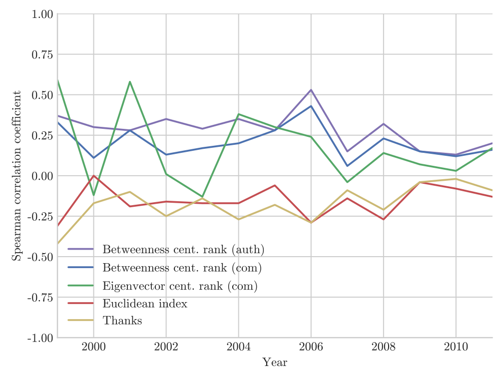
(b) Number of Thanks



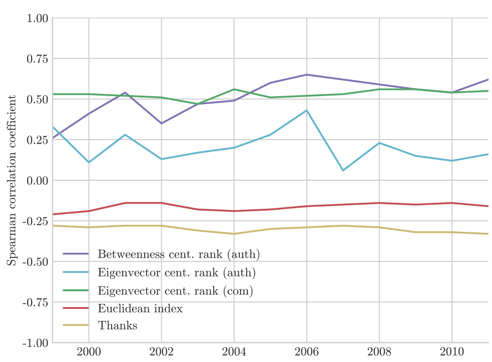
(c) Betweenness centrality rank (co-author network)



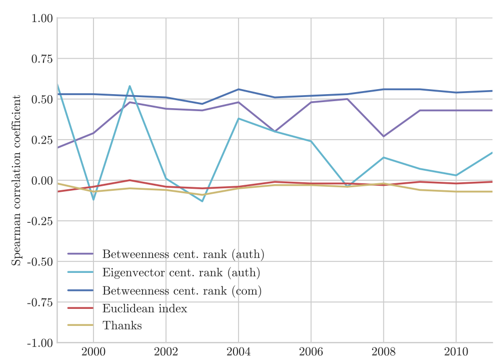
(d) Eigenvector centrality rank (co-author network)



(e) Betweenness centrality rank (commenter network)



(f) Eigenvector centrality rank (commenter network)



Notes: Figures depict Spearman rank correlation coefficients over time for various variables. For each year t, the sample includes publications in six financial economics journals published in t, $t-1$ and $t-2$. *Euclidean index* is the researcher's Euclidean index of citations according to equation (2.1). *Thanks* is the number of papers that acknowledge this researcher. *Eigenvector cent. rank* is the researcher's eigenvector centrality in the giant component of the network of informal (com) and formal (auth) collaboration according to equation (2.4). *Betweenness cent. rank* is the researcher's betweenness centrality in the giant component of the network of informal (com) and formal (auth) collaboration according to equation (2.5).

Table 2.11 highlights that often thanked scientists are not necessarily the most relevant for the flow of information, nor the most connected authors. This is corroborated by the correlation coefficients depicted in figure 2.9. It shows Spearman correlations over time between all centralities, the Euclidean index of citations and the number of thanks. Subfigure 2.9a confirms that being thanked often and being a very prolific academic is not the same, as the Spearman rank correlation coefficient never surpasses 0.3. A low correlation also exists between being thanked often and being central in the co-author network (subfigure 2.9b). All correlation coefficients are very stable over time.

2.4 Discussion

2.4.1 Strategic Acknowledging

An important question to address is whether authors use their acknowledgements (predominantly) strategically. In this case acknowledgements would not necessarily reflect informal collaboration between authors and acknowledged commenters. Following Hamermesh (1992, p. 171), we define strategic acknowledging as an author's attempt to influence an editor in her choice of referees by acknowledging "someone who has not seen the paper, as a talisman against that person being chosen." Even though there might be conflicting views,[29] the general assumption seems to be that editors do not pick already acknowledged commenters. According to this assumption, authors would want to thank someone that has a reputation of being a tough referee.[30] Irrespective of whether editors actually behave according to this view, acknowledging someone who has not actually given a comment carries a high reputation risk. If that person learns about it (e.g. during the review process), it will reflect badly on the author.[31]

While we have no doubt that some authors use the acknowledgement section of their pa-

[29]Other authors have indicated they believe that editors would prefer to pick someone who is acknowledged.

[30]This strategy is summarized in *"Cite your friends, acknowledge your foes."* Editors of various journals, however, have indicated to us that they seldom exclude a potential referee simply because this person is acknowledged. Also, not all editors explicitly look into the acknowledgment section when selecting a referee.

[31]Because of this risk, Hamermesh (1992, p. 171) writes in his "Guide to Professional Etiquette": *"DON'T PLAY THESE GAMES - the gains are not worth the potential cost of being caught"* (emphasis in the original).

per as a signaling device (e.g. to influence the editor), we observe a number of stylized facts which indicate that this is not a systemic phenomenon. First, authors do not put the names of senior and prominent researchers first in the acknowledgment section. Rather, authors usually order the names of commenters alphabetically. Second, the list of commenters is not always first in the acknowledgement section (which in total may well span more than 10 lines). Seminars, conferences, research assistance or funding are listed before commenters. Third, more than half of all papers acknowledge individuals that no other publication acknowledges. This speaks against the view that all acknowledged commenters are put down for strategic reasons, as there is little signaling value in thanking researchers that are relatively unknown to Financial Economics as a field. The low correlation between being frequently acknowledged and being prolific reported earlier also shows that authors do not predominantly acknowledge prolific authors.

Another form of intentional acknowledging could still exist in our data. Authors could strategically seek advice from senior and well-known researchers. This variant of strategic acknowledging, however, is precisely what we want to capture. Authors identify scholars that they think might be of help for an ongoing research project and with whom they subsequently try to collaborate. For our analysis it is not relevant why scholars discuss with each other, as long as they actually collaborate.

2.4.2 Relation to Other Studies in the Field and External Validity of our Data

In order to establish more confidence in our novel data, we report replication for major studies in the field. We replicate two studies with our data, Laband and Tollison (2000) and Brown (2005). Both estimate the impact of informal intellectual collaboration on the number of citations of published papers. We are able to replicate the results and sometimes find stronger correlations. Thus our data are akin to the data used in the existing literature.

Laband and Tollison (2000) use 251 featured articles published in the Review of Economics and Statistics during the years 1976-1980. They estimate the effect of the number of acknowledged commenters to explain the number of citations the paper receives over the following six years. They control for the cumulative stock of citations from the previous five years for all authors, as well as the number of pages. They show that the number of commenters is statistically significantly and positively associated with the paper's citation count. In alterations to the model they add the commenters' joint citation stock over the previous five years, and the count of commenters that are a) not at the same department as the authors, b) on one of the author's dissertation committee, c) at the same department as one of the authors, and d) commenter not belonging to one of the previous groups. Columns (1) through (3) in table 2.12 replicate model (1) through (3) of table 4 of Laband and Tollison (2000). While models (1) and (2) are similar, in model (3) we find the number of acknowledged commenters still to be statistically and economically significant after controlling for their caliber.

Brown (2005) uses a negative binomial regression similar to ours and a sample of 256 papers published in The Accounting Review, the Journal of Accounting Research, and the Journal of Accounting and Economics during 2000-2002. The dependent variable to measure publication success is the number of citations since publication according to the Social Science Citation Index. His main explanatory variables are the number of commenters, the number of conferences, and the number of seminars. Brown (2005) controls for the number of pages, the number of authors, whether the paper was highly downloaded from SSRN, and also uses journal- and time-fixed effects. He finds that only seminars have a statistically significant and positive impact on citation count. Estimating the impact of acceptance probability on the journal he edited—The Accounting Review—he finds that all forms of informal intellectual collaboration matter. Column (4) table 2.12 reports replicates of estimates presented in Brown (2005, Table 8C), with the difference that we do not control for the number of downloads from SSRN. However, for our sample we find a statistically significant relationship between the number of commenters and citation count, even after controlling for the number of acknowledged seminars and conferences.

Table 2.12: Regression results replicating Laband and Tollison (2000) and Brown (2005)

	Six-year citations			Total citations
	OLS			*negative binomial*
	(1)	(2)	(3)	(4)
Authors' 5-year cites	0.002*** (0.0001)	0.002*** (0.0001)	0.002*** (0.0001)	
No. of pages	0.756*** (0.057)	0.891*** (0.065)	0.799*** (0.066)	
No. of authors				0.171*** (0.015)
No. of commenters	1.181*** (0.088)		0.819*** (0.126)	0.021*** (0.002)
Commenters' 5-year cites		0.0003*** (0.00003)	0.0002*** (0.00003)	
No. of seminars				0.013*** (0.004)
No. of conferences				0.0003 (0.008)
Constant	0.984 (1.431)	2.866 (1.792)	−0.212 (1.846)	4.092*** (0.067)
Journal-fixed effects	*No*	*No*	*No*	*Yes*
Publication year-fixed effects	*No*	*No*	*No*	*Yes*
N	6,356	5,291	5,291	6,370
R^2	0.141	0.132	0.139	
Adjusted R^2	0.141	0.132	0.138	
Log Likelihood				-33,448.640
Residual Std. Error	43.815 (df = 6352)	46.796 (df = 5287)	46.614 (df = 5286)	

Notes: ***, ** and * indicate statistical significance to the 1, 5 and 10 percent level, respectively. Column 1 through 3 replicate models 1 through 3 of Laband and Tollison (2000, Table 4). Column 3 replicates Panel B of Brown (2005, Table 8), with a slightly different variable definition and without the SSRN control variable. Reported coefficients in column 4 are marginal effects and show the per cent increase in the citation count in response to a 1 unit increase in the independent variable, holding all variables at their mean and setting binary variables to 0. *Authors' 5-year cites* is the sum of individual citation stocks (according to Scopus) for all authors for the five years prior to the publication year. *No. of pages* and *No. of pages* is the count of pages and authors, respectively. *No. of commenters* s the count of all commenters acknowledged for concept-related input (excluding editors). *Commenters' 5-year cites* is the sum of individual citation stocks (according to Scopus) for all commenters acknowledged for concept-related input (excluding editors) for the five years prior to the publication year. *No. of seminars* and *No. of conferences* is the count of seminars resp. conferences acknowledged in the papers' acknowledgment section.

2.5 Conclusion

This chapter highlights the importance of informal collaboration among financial economists. The nascent literature on collaboration in academia focuses mostly on formal collaboration, i.e. co-authorship, but most collaboration among researchers in financial economics is *in*formal, in the form of commentary and feedback. We use a novel and unique dataset obtained from the acknowledgement sections of $6,401$ research papers published in six scholarly journals in financial economics between 1997 and 2011. Using this data set, we present a number of novel results about informal collaboration, focusing on the social network formed when researchers collaborate informally. Specifically, we show that a researcher's position in this network of informal collaboration–measured as her eigenvector centrality, a well-established network of influence in netowrks–correlates with her future productivity and also with the academic impact of papers she comments on. This is particularly useful information in case of young researchers who have no or only few publications when applying for a job.

Once we have established that the network of informal collaboration is more than just a random collection of bilateral collaborations and that a researcher's position actually contains useful information, we study this network in more detail. We show that researchers with a high betweenness centrality in the network of informal collaboration, i.e. researchers who connect disparate research communities, are less likely to publish in top finance journals, but if they do, their papers receive more citations than the average paper published in the same journal. We then show that the number of researchers who are involved in producing research in financial economics is roughly twice as large as the number of authors who publish themselves. This results in the network of informal collaboration being much more densely connected than the network of formal collaboration.

Finally, we turn to an analysis of what makes a researcher more central in the network of informal collaboration and then provide a list of the most central financial economists from 1997 to 2011. We show that more senior researchers are less eigenvector central, while more prolific researchers have a higher betweenness centrality and thus are more likely to connect otherwise disparate research communities. We provide the list of the most central researchers to build intuition for our results and to highlight those researchers who are, in the terminology of Oettl (2012), more helpful.

Informal collaboration is vital in other disciplines as well (Cronin, 1995). Studying informal collaboration, as well as the flow of information it reveals, can help understand differences in (perceived) quality among journals, papers, authors and affiliations. One aspect of this chapter, and the entire dissertation, is to take on a network perspective, and to show how networks of formal and informal collaboration differ. The growing literature on “knowledge networks”, reviewed by Phelps et al. (2012), highlights this nicely. We demonstrate the informational content of networks of informal collaboration and show how they differ from co-author networks. The topology of these networks is of interest, because it affects the speed of learning and the diffusion of information (Alatas et al., 2016).

The analysis of collaboration adds new insights into the division of labor in academic teams. There is a wide range of activities that are necessary for scientific innovation (Haeussler and Sauermann, 2016). Not all of these need to be performed by formal collaboration: Authors can extend the team to outsource activities that do not justify co-authorship alone. Put differently, a group of researchers produces an academic paper, but this research group may be larger than the actual authors (Ponomariov and Boardman, 2016).

This chapter does not aim to causally identify what determines patterns of formal or informal collaboration. Rather, it highlights a number of interesting observations based on our novel data. These observations give rise to an array of interesting questions, though. What is the underlying mechanism for the correlation between eigenvector centrality and a researchers’ future productivity? What is the mechanism through which researchers improve the academic impact of a research paper with their commentary? What incentivizes researchers to comment on each others work, given that we do not find strong evidence for reciprocity? What explains the age effects in the provision of commentary? Why are female researchers statistically significantly less often acknowledged at same levels of academic productivity and experience? What caused the statistical malus for female researchers in eigenvector centrality to disappear around 2006? When do researchers decide to collaborate informally? When do they collaborate formally? Is there a transition from informal to formal collaboration, or vice versa? These questions are relevant to the profession and our study is a first step towards a broader understanding of the role informal collaboration plays in the creation of knowledge.

Chapter 3

Informal Collaboration with Central Colleagues

3.1 Introduction

In this chapter, we show that intellectual collaboration positively influences research output through information spillover and complementarities in research efforts. A researcher absorbs more information about the work of other researchers when she exerts more effort on her research activities–she becomes more knowledgeable. Collaborating with a more knowledgeable researcher implies access to more accumulated knowledge and hence greater benefits from information spillover.

Several authors have taken steps to explain the sources of increase in collaboration (Hudson, 1996; Laband and Tollison, 2000; Goyal et al., 2006) and show that intellectual collaboration among researchers has a positive impact on their productivity (Wuchty et al., 2007; Waldinger, 2012; Ductor et al., 2014; Ductor, 2015). By intellectual collaboration we mean both formal collaboration in the form of co-authorship, and informal collaboration in form of advice and feedback on an ongoing project.[1] Besides its positive impact on productivity, understanding how

[1] See Laband and Tollison (2000) for a similar characterization of the term. Economists are prominently advised to seek feedback and commentary from their colleagues on a research article prior to publication (Green et al.,

the social network of intellectual collaboration affects the outcome of research projects is also relevant in light of recent trends in academia, such as the increasing competition for space in top scholarly journals (Card and DellaVigna, 2013), the increasing time lag until publication (Ellison, 2002), the increasing duration of education of researchers (Jones, 2009). The mechanisms that drive the positive relationship between collaboration and research output are, however, not clearly understood.

Our notion of intellectual collaboration is akin to the notion of absorptive capacity in the literature on R&D investment by firms. This notion was first developed by Cohen and Levinthal (1989) who show that not only does R&D generate new information, it also enhances the firm's ability to assimilate and exploit existing information. The latter leads to complementarities in R&D investments through information spillovers. We argue that the same principle applies to intellectual collaboration in fundamental scientific research.

We develop a simple model of intellectual collaboration in academia that incorporates the notion of strategic complementarity in efforts. In our model, academics collaborate in a social network of intellectual collaboration. They derive utility from the net output from research activities. That is, the total quality and quantity of their output less the opportunity cost of not working, i.e. enjoying leisure time. We control for the quality of research output by normalizing effort by the number of projects a researcher undertakes in a given period of time. To model strategic complementarity in efforts, we assume that the marginal output depends positively on a researcher's intrinsic characteristics, and the effort exerted by her collaborators.

The existence of a term capturing complementarity in efforts implies that equilibrium allocation of effort exercised by an academic is a function of the academic's Bonacich network centrality (Ballester et al., 2006). The aggregate equilibrium effort and hence total output increases with connectivity. Overall, in equilibrium, researchers with the highest centrality are also those with the highest effort. This implies that collaborating with central colleagues leads to higher level (quantity and quality) of individual output. At the project level, it follows that collaborating with central colleagues leads to a high quality project, and hence higher scientific

2002).

impact, measured by the number of citations a research article receives.

We test this hypothesis in the academic field of financial economics, which has the advantage of being a large and a relatively homogeneous sub-field of economics. Our main data source, which was introduced in the previous chapter 2 ,are the title pages of 6,401 full research articles from six scholarly journals in financial economics published between 1997 and 2011. From these articles we manually construct a novel and unique data set consisting of all authors and commenters acknowledged in a research article. Each network consists of three consecutive years of publications and connects authors to acknowledged commenters and vice-verse. Links in the network are directed and weighted. Weights correspond to the frequency of interaction and by the productivity of the target node, where productivity is defined as Euclidean index of citations divided by number of years since first publication (experience).

To overcome endogeneity inherent to social networks (Graham, 2015), we compute centrality scores in a network where deceased authors have been removed with a network where they have not. The change between two centrality scores is then purely due to the exogenous removal of deceases authors.

We assess the impact of commenter centrality change on citation count in a negative binomial regression and find significant support for our hypothesis. We find, in our main result, that an increase in Bonacich centrality by 2% of the average most central commenter is associated with an increase by about 1 citation for the average article–after addressing inherent network endogeneity and after controlling for an extensive set of author, discussant, and bibliometric characteristics. Our results are robust to a variety of alternative specifications, including different attenuation factors in the computation of the Boncich centrality, different centralities and different network specifications.

The chapter contributes to various strands of the literature. First, it contributes to the literature that studies the impact of intellectual collaboration on research output. Hollis (2001) finds that academic teamwork has positive effects on quality, length and the number of published articles. In addition, he finds that increasing co-authorship in the past (conditioning on current average co-authorship and the lifetime number of articles) increases the likelihood that an academic is prolific today. Hollis (2001) attributes this to learning that occurs in the collaborative process. A similar result is reflected upon in Ductor et al. (2014), who find that an early career academic's network of formal intellectual collaboration helps predict their future productivity. Azoulay et al. (2010) provide evidence of information spillovers and find that the loss of a superstar academic leads to a lasting $5-8\%$ average decline in the quality-adjusted publication rates of co-authors. Waldinger (2012) shows that professors' productivity drops after they lose a co-author due to dismissal. The chapter contributes to this debate by showing that the positive relationship between intellectual collaboration and research output can be explained by the relationship between information spillover and complementarities in efforts.[2]

Secondly, the chapter contributes to the theoretical literature on collaboration networks in academia and R&D (e.g. Goyal and Moraga-González (2001), König et al. (2014) and Hsieh et al. (2018)). Goyal and Moraga-González (2001) study collaboration in R&D in the presence of externalities and show that under strong market rivalry, R&D effort declines with the level of collaborative activity. In the absence of firm rivalry, however, R&D effort increases with the level of collaborative activity. Under a similar set up König et al. (2014) show that Nash equilibrium output of firms is proportional to their Katz-Bonacich centrality in R&D network. The respective optimal output choice depends on the competition intensity the firm faces in the product market. Hsieh et al. (2018) study a model similar to ours and provide conditions for existence and uniqueness of an interior equilibrium. We differ from their model along an important line: we control for the quality of research output, and we show how equilibrium efforts depend on the distribution of individual productivities. More generally, our model contributes to the literature of network games that identify the role of individual centrality in the network on their equilibrium behavior (e.g. Ballester et al. (2006), Bramoullé et al. (2014) and for a comprehensive survey of the literature, see Jackson and Zenou (2015)).

Finally, the chapter complements a small number of papers that investigate the relationship between informal intellectual collaboration and the research process: Laband and Tollison

[2]Management literature also studies whether collaboration can improve the quality and economic value of knowledge produced. Examples include Singh and Fleming (2010) who show that collaboration reduces the probability of very poor outcomes while simultaneously increasing the probability of extremely successful outcomes. Girotra et al. (2010) find that hybrid team structures, in which individuals first work alone then work together, are able to generate more ideas, generate better ideas, and to better discern the quality of ideas they generate.

(2000) focus on the social aspect of informal intellectual collaboration in Economics and Biology. They find that a higher number of commenters is associated with a higher citation count over seven years. But the benefit increases in the citation count–the so-called "caliber"–of the commenter (in our nomenclature: how prolific a commenter is). Unlike Laband and Tollison (2000), Brown (2005) also includes other forms of informal intellectual collaboration, such as seminar presentations. He finds that the number of acknowledged seminars is more relevant for citation count than the number of commenters. The same is true for the acceptance probability at prestigious Accounting journals. But neither of these studies, which we have replicated in section 2.4.2, take into account the network structure of the social network prevalent in Economics and its sub-fields. Oettl (2012) takes a near-network perspective by estimating the malus co-authors of very eminent life scientists experience when these eminent scientists die. The former co-authors' drop in quality-adjusted research-output amounts to 20% in this measure. Interestingly, the most important channel in Oettl (2012) is not formal, but informal intellectual collaboration. For this reason, Oettl (2012) terms this dimension "helpfulness". Our main contribution to this literature, besides being the first to study informal intellectual collaboration in an entire academic field, is that we address endogeneity through our empirical setup. Furthermore, our theoretical model provides a conceptual framework that allows us to empirically disentangle information spillovers from strategic complementarities.

3.2 A Simple Model of Collaboration in Academia

The goal of this section is to provide a simple model framework for intellectual collaboration with complementarity in efforts. We model a set $N = \{1, \cdots, i, \cdots, n\}$ of researchers who engage in research to increase the quantity and quality of their output Y_i. The production process involves individual effort as well as effort of others through intellectual collaboration; that is both formal collaboration in the form of co-authorship and informal collaboration in form of receiving feedback from other researchers. In addition to responding to a request from a colleague to provide feedback on an ongoing project, informal collaboration also includes being a discussant at a conference, as well as inter-departmental collaboration through research seminars.

Intellectual collaboration contributes to output Y_i by exposing an ongoing research project to other scholars who potentially give feedback. Such feedback not only improves the quality of the current project but may also give insights to ideas for new projects, and hence even increase the total number of projects.

We assume that total output is linearly decomposable into a direct contribution and an indirect contribution from complementarities. As discussed in Section 3.1, complementarity in efforts contributes to research output through information spillover. At the level of an individual researcher, information spillovers with co-authors forces one to increase their effort to be able to assimilate the knowledge and techniques of co-authors. The knowledge and techniques learned not only improve the quality of the paper in progress but will also be used as input to future projects. At the level of the paper, the same notion applies with contributions from informal collaborators. A direct consequence of such interactions is an increase in the overall productivity of a researcher; that is, the marginal output. As discussed in section 3.1, several papers have documented strong evidence of such externalities in collaboration. For example Ductor (2015) finds a positive effect of intellectual collaboration on individual productivity. Wuchty et al. (2007) find an increasing dominance of teams in knowledge production, and show that teams produce exceptionally high impact research compared to solo authored work. They also find that research produced in teams tends to be more cited than that by individual authors.

Formally, let e_i denote the effort of researcher i and $\mathbf{e}$ the vector of efforts. Let G be a collaboration network among researchers. With slight abuse of notation, we also write G for the adjacency matrix of G. That is, each element g_{ij} of G is defined in such a way that $g_{ij} = g_{ji} = 1$ if researcher i collaborates with j and zero otherwise. We assume that the links are undirected. For the case of informal collaboration, this results from the idea that in the process of giving feedback, a researcher also learns about the methods and results of someone else's work. Ideally, the value of g_{ij} should be different, depending on whether j is a co-author or informal collaborator. This would clearly distinguish the level of information spillover between the two types of interactions. Here, for the sake of simplicity, we assume that the level of information spillover is identical in both cases. Moreover, there is evidence suggesting that informal collaboration is just as effective as formal collaboration in generating ideas for research. For example

Colander (1989, p. 146) concludes his survey that "[m]uch if not most of the debate and discussion about economic ideas take place at the pre-working paper, workshop and working paper stages.", and Ductor et al. (2014, p. 937) argue in a study on productivity patterns among co-authors that *"a researcher who is close to more productive researchers may have early access to new ideas"*. Nevertheless, in our empirical analysis below, we consider both the network of only informal and a combined network of formal and informal collaboration.

For each i, let N_i denote the set of first-order neighbors. That is, the set of all agents who directly collaborate with i. Let n_i be the cardinality of N_i. To control for quality of each output, let p_i be the total number of projects i is involved in. Assuming that i allocates her effort equally across the p_i projects, effort exerted on each project is $\frac{e_i}{p_i}$.[3] The total output $Y_i(G,\mathbf{e})$ to i for engaging in collaboration network G, given effort configuration $\mathbf{e}$ is then:[4]

$$Y_i(G,\mathbf{e}) = e_i t_i + \sum_{j\in N_i} g_{ij}\frac{e_j t_j}{p_j} + \alpha \sum_{j\in N_i} g_{ij} e_j e_i \quad , \tag{3.1}$$

where t_i, the type of i, is individual productivity of i, which is determined by i's intrinsic characteristics such technical skills and seniority in the field. The first two terms on the right hand side of (3.1) capture the direct contribution of i and her set of collaborators. The third term captures complementarity in efforts, where α is a parameter capturing the contributive strength of such externalities. Following the above discussion, we see from (3.1) that the overall productivity of i is $t_i + \sum_{j\in N_i} g_{ij} e_j$. The second term captures the notion that the effort of i's collaborators positively influence her productivity. It is related to the idea of Jackson and Wolinsky (1996) and is a kind of a congestion externality from collaboration: The larger the number of project of i's coauthors, the less the effort they devote to i's projects.

Associated with effort level e_i, is an opportunity cost $c_i(e_i)$. We let $c_i(e_i)$ assume a quadratic form with parameter β identical for all agents for simplicity, i.e. $c_i(e_i) = \frac{1}{2}\beta e_i^2$. The utility $U_i(G,\mathbf{e})$ that i derives from network G while exerting effort e_i is then the net output $Y_i(G,\mathbf{e})$ −

[3] This assumption simplifies our computations and comes at no loss of generality: a researcher will spend more time on projects where she is a co-author, but this could be easily adjusted for by using a simple multiplicative factor.

[4] Note that the first term of the right hand side of (3.1) results from summing $\frac{e_i}{p_i}$ over all projects p_i.

$c_i(e_i)$. That is:

$$U_i(G,\mathbf{e}) = e_i t_i + \sum_{j\in N_i} g_{ij}\frac{e_j t_j}{p_j} + \alpha \sum_{j\in N_i} g_{ij} e_j e_i - \frac{1}{2}\beta e_i^2. \tag{3.2}$$

The model specification in (3.2) has similarities with those in the literature of network games (e.g. Ballester et al. (2006), Bramoullé and Kranton (2007) and Bramoullé et al. (2014)). The similarity is in the existence of externalities. The difference arises in the nature of the production process we model, which involves group production as captured by the first two terms of (3.2), and that we control for the quality of output. We characterize equilibrium properties of the game, and in particular conditions for existence and uniqueness of an interior equilibrium solution. We then use the resulting equilibrium efforts in an empirical study where we test for the existence of complementarities and information spillover in collaboration. We achieve this by assuming that equilibrium effort of a collaborator has a positive impact on the quality and hence impact of a researcher's output. We proxy the impact of research output by the citation count it receives. In particular, we estimate models of the following kind:

$$Citation_i = F(\mathbf{t},\mathbf{p},\mathbf{e}^*,\mathbf{c}), \tag{3.3}$$

where we write $F(a)$ to imply a function of a. The independent variables are $\mathbf{t}$ which is a vector of types, $\mathbf{p}$ the vector of number of projects, $\mathbf{e}^*$ which is equilibrium outcome and $\mathbf{c}$, a vector of control variables. We use citations of research output as a measure of its impact.

3.3 Equilibrium Properties

Given (3.2), each researcher chooses an optimal level of effort e_i^*, where the first order condition is:

$$\beta e_i^* = t_i + \alpha \sum_{j \in N_i} g_{ij} e_j^* \quad \text{for each } i \in N. \tag{3.4}$$

It is well known (e.g. Ballester et al. (2006)) that equilibrium levels of effort in a game set up such as (3.2) depend on the Bonacich centrality of the underlying network of interactions, defined as follows. Given a scalar $\lambda \geq 0$ and a network G, let a matrix $M(G,\lambda)$ be defined as

$$M(G,\lambda) = (I - \lambda G)^{-1} = \sum_{k=0}^{+\infty} \lambda^k G^k$$

where I is the identity matrix. Let $\mathbf{1}$ be a column vector of ones. For an $n \times n$-square matrix G and a scalar λ such that $M(G,\lambda)$ is well defined, the vector of centralities of parameter λ in G is:

$$\mathbf{b}(G,\lambda) = (I - \lambda G)^{-1}\mathbf{1}$$

The Bonacich centrality of node i is $b_i(G,\lambda) = \sum_{j=1}^{n} m_{ij}(G,\lambda)$, and counts the total number of paths in G starting from i. For a vector $\mathbf{t}$ of t_i's, we define a corresponding vector of centralities $\mathbf{b}(G,\lambda,\mathbf{t})$ as

$$\mathbf{b}(G,\lambda,\mathbf{t}) = (I - \lambda G)^{-1}\mathbf{t}$$

Proposition 1. *Let $\mu_1(G)$ be the maximum eigenvalue of G, The game with payoffs in (3.2) has a unique interior equilibrium whenever $\beta > \alpha\mu_1(G)$. The respective equilibrium configuration is* *given by*

$$\mathbf{e}^* = \frac{1}{\beta}\mathbf{b}(G, \alpha_\beta, \mathbf{t}) \tag{3.5}$$

where $\alpha_\beta = \frac{\alpha}{\beta}$.

Proof. See Appendix 3.A □

Proposition 1 provides a characterization of equilibrium efforts. Equilibrium efforts are a function of Bonacich centralities of the network, a property that is well known in the literature of network games. Overall, scientists who collaborate with many other scientist will have a high Bonacich centrality. Similarly, collaborating with scientists who are themselves highly central, increases ones centrality even more. As a corollary, increasing the overall level of connectivity in the network in turn increases the overall equilibrium effort (Ballester et al., 2006; Bramoullé et al., 2014).

The component of the centrality measure in (3.5) that is specific to our model is its dependence on the distribution of types of researchers. Researchers with higher individual productivity exert higher effort at equilibrium, as observable from (3.4). On an aggregate level, researcher with higher productivity have a positive impact on aggregate effort. That is, since they tend to exert a higher effort, other researcher who directly collaborate with them would also exert higher effort due to effect of complementarities, and so will the collaborators of collaborators, and so on. This effect is observable from (3.5) and the fact that $\mathbf{b}(G, \alpha_\beta, \mathbf{t}) = \left(I - \frac{\alpha}{\beta}G\right)^{-1}\mathbf{t}$, where the term $\left(I - \frac{\alpha}{\beta}G\right)^{-1}$ acts as a multiplier on the types. This effect is strongest if individuals with the highest productivity are also the most central.

There are two main conclusions that follow from our stylized model of intellectual collaboration. First, high connectivity, and hence high intensity of collaboration among researchers leads to higher aggregate equilibrium outcome. Second, for a given network of collaboration and distribution of types, the optimal aggregate equilibrium outcome is obtained in a set up where the most central individuals are also those with the highest individual productivity. In the following, we bring these results to the data and test them empirically.

3.4 Variable construction

3.4.1 Centrality in the Social Network of Informal Collaboration

To construct networks of informal collaboration, we use the data introduced in section 2.2: Hand-collected acknowledgments from 6,401 articles published between 1997 and 2011 in six scholarly journals in financial economics.

We connect two researchers whenever one acknowledged the other as a commenter on a published research article in our dataset. Links between these researchers are directed but always have a parallel link in the opposite direction. This is based on the notion that knowledge spillover occurs in both directions: The author tells the commenter, sometimes in great detail, about the article, which is valuable to the commenter. For example she can use the results to build her own research on it before it is published. Spillovers from the commenter to the author occurs in the form of feedback, which in turn not only improves the quality of the author's current work but may also provide ideas for future research. Links are weighted by the frequency of interaction between collaborators.

For each year t we construct the network using the publications published in t, as well as in the two previous years, $t-1$ and $t-2$. Formally, let A_t be the set of articles published in years $\{t, t-1, t-2\}$. To each article $a \in \{A_t\}$, there is a non-empty set of authors κ_a and a not necessarily non-empty set of commenters ι_a. Every author $i \in \kappa_a$ and every commenter $k \in \iota_a$ is part of the set of nodes that either authored or acknowledged in the set of articles A_t. The resulting network G is weighted in such a way that for each pair i, j, g_{ij} increases by $1/|\kappa_a|$ if author i acknowledges commenter j on article a. If the commenter has been acknowledged once on an article written by two authors, the weight of each of the two ties would be $1/2$. If one of the authors acknowledges this commenter on another solo-article, the tie increases to $3/2$, reflecting a deeper relationship between the two. The adjacency matrix G is symmetric as acknowledgements are undirected, with the diagonal elements being equal to 0. This weighting scheme also corrects for misreporting. That is, when a research article consists of many authors, it is not clear which author spoke to which commenter. This is thus accounted for by the weights of $1/|\kappa_a|$.

In real world networks it is common that nodes are not connected, not even indirectly. Formally, two nodes belong to the same network component $\updownarrow$ if there exists an alternating sequence of nodes and ties, called a path, between them. There can be as many components as there are nodes if all nodes are isolated. The size of a component is the number of nodes (i.e. the number of academics) it contains. The component containing the most nodes is called the giant component. For the regression, we compute and use the Bonacich centrality for all nodes in the giant component and omit the other components because centralities across components are not comparable: If a node i belongs to a small network component, all other nodes are fairly close. In contrast, a node in a large network might have a potentially much smaller centrality because many other nodes are far away.

We construct twelve networks for all t between 2000 and 2011. As the number of articles increased over time, the network increase by size, too: The 2000 network is generated from 873 articles published in either 1998, 1999 or 2000 and consists of 3,286 distinct researchers. This compares to the 2011 network, which connects 7,028 researchers that have collaborated on 1,889 articles. The giant component captures between 95% and 98% of the network, implying that the error that we make from investigating the giant component only is relatively small.

In (3.5) of Proposition 1, we show that equilibrium efforts $\mathbf{e}^*$, are equal to the Bonacich centralities $\mathbf{b}(G,\alpha,\mathbf{t})$ of network G weighted by the vector $\mathbf{t}$ of individual productivities.[5] That is

$$\mathbf{e}^* = \mathbf{b}(G,\alpha,\mathbf{t}) = (I-\alpha G)^{-1}\,\mathbf{t} = \sum_{k=0}^{+\infty} \alpha^k G^k \mathbf{t} \tag{3.6}$$

The parameter α is generally referred in the literature as *attenuation factor*. It discounts the distance between agents that are not collaborating directly. The smaller α, the less impact

[5] Without loss of generality, we take $\beta = 1$ such that $\alpha_\beta = \alpha$, and $\mathbf{e}^* = \mathbf{b}(G,\alpha,\mathbf{t})$.

distant agents have on an agent's equilibrium effort. In our model, α captures the contribution of complementarities in collaboration to ones research output. Proposition 1 provides equilibrium conditions; that is $\alpha < \frac{1}{\mu_1(G)}$, where $\mu_1(G)$ is the leading eigenvalue of the weighted network G. In the regression, we use $\alpha = \frac{0.85}{\mu_1(G)}$. In section 3.6.1 we test different attenuation factors and observe that coefficients peak in size and statistical significance for an attenuation factor of around 0.80.

3.4.2 Author metrics

As before, we proxy individual productivity by the quantity and quality of their overall output using the Euclidean index of citations (Perry and Reny, 2016). For every year, the Euclidean index is the square root of the sum of the squared number of citations to each individual published article (2.1). The alternative measures would be the total number of publications or the total citation count normalized by the number of years of experience. The Euclidean index however takes into account both measures making it a more accurate measure. We use data obtained from Scopus to compute the Euclidean index, which is hence available for 11,718.

The number of projects each researcher is involved in each year is the number of publications in the current year and the next year. To account for shared work, each project is divided by the number of co-authors. For example for someone that published one single authored article in 2005, and one co-authored article in 2006, the number of projects in 2003 would be 0, in 2004 1, in 2005 1.5, and in 2006 0.5.

3.4.3 Deceased authors

We strive to identify the effect of Bonacich centrality by comparing centralities computed in a network in which deceased scientists are removed with a centralities computed in a counterfactual network. Those two networks differ in that deceased authors were removed from the first. If the difference in centrality scores between these two networks is positive then the commenter's network position increased.

Our main source of deaths is the IDEAS database of Research Papers in Economics (RePEc) repository hosted by the Federal Reserve Bank of St. Louis.[6] We augment this list with information from the English-language version of Wikipedia, various university websites and acknowledgements itself. We find 61 researchers that deceased within two years of their last recorded activity in our yearly networks (i.e. provision of comment or publication). Table 3.B.1 in the appendix lists them by date of death.

3.5 Identifying the Impact of Informal Collaboration on Publication Impact

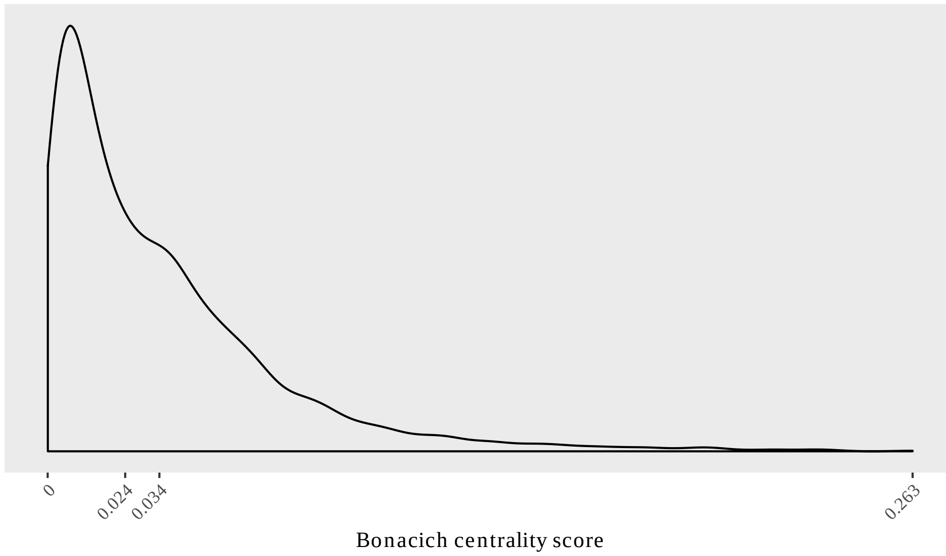
The unit of observation is a published research article. The relevant social network from which we compute centralities is always inferred the network of the previous year. Hence the author's own network status does not influence the positions of the commenters.[7] We study all research article published in a set of six major finance journals published between 1998 and 2011 that have acknowledged at least one commenter who is in the giant component of the social network of informal intellectual collaboration ($N=1,415$). These journals are The Journal of Finance, The Review of Financial Studies, the Journal of Financial Economics, the Journal of Financial Intermediation, the Journal of Money, Credit, & Banking, and the Journal of Banking and Finance.

To aggregate over all acknowledged commenters, we look at the most central commenter in our main regression. That is, we ask what if the authors could only receive comments from the most central commenter? To estimate robustness of the results, we also estimate our main specification with the second-most central commenter in a sample of publications acknowledging at least three commenters by name. To give an idea of the distribution of the most Bonacich central commenter's centrality scores, figure 3.1 depicts median, mean and maximum. The distribution ranges between 0 and 0.455, with a median of 0.006 and a mean of 0.024.

[6]The full list of author profiles with confirmed deaths is available at https://ideas.repec.org/i/erip.html.

[7]A possible exception is when authors and commenters have collaborated in previous years.

Figure 3.1: Distribution of Bonacich centrality of most central commenter.



Notes: Graph shows distribution of Bonachich centrality (equation (3.6)) of the most central commenter of an article, in the giant component of the network of informal collaboration in the year before publication without deceased authors removed. Abcissa depicts minimum, median, mean and maximum.

Our regression estimation obtains as:

$$\begin{aligned} Citations = \alpha_0 + \alpha_1 \cdot ArticleCharacteristics \\ + \alpha_{\mathbf{1}} \mathbf{D}_{journal} + \alpha_{\mathbf{2}} \mathbf{D}_{year} \\ + \beta_0 \cdot AuthorCharacteristics + \beta_1 \cdot BestCommenterCharacteristics \\ + \beta_2 \cdot BestCommenterCentralityChange \end{aligned} \tag{3.7}$$

where *ArticleCharacteristics* is a vector of article-specific variables that contain the number of authors, the number of pages, number of authors, number of acknowledged seminars and number of acknowledged conferences. $\mathbf{D}_{journal}$ is a vector of dummy variables that captures journal fixed effects such as editor skills and preferences or management policies, while $\mathbf{D}_{year}$ represents publication year fixed effects. *AuthorCharacteristics* and *BestCommenterCharacteristics* are vectors controlling for characteristics of all authors and the best commenter. For authors this includes the sum of the Euclidean indices of citations of all authors, the sum of other author-normalized projects of all authors, the sum of experience of all authors and finally the sum of squared experience of all authors. Squared experience are intended to capture non-monotonic age effects. Since we also control for the number of authors of a research article we indeed estimate the mean effect of author characteristics. *BestCommenterCharacteristics* include the number of projects. We do not include productivity measures because this is included in the computation of the Bonacich centrality. All these variables were counted in the year before publication.

Our main contribution is to introduce *BestCommenterCentralityChange*, the change in centrality according to Bonacich centrality as defined in (3.6) in the social network of informal intellectual collaboration of the most central commenter. These changes, while random, are not correlated with centrality scores.

The dependent variable is the count of citations since publication. The distribution of citations is skewed, discrete and non-negative. Therefore we estimate a negative binomial regression model (Mullahy, 1986). Being a generalized linear model, the parameters are evaluated at sample mean, i.e. holding all variables fixed at their mean values. To ease interpretation, we compute and present marginal effects, i.e. the coefficients we present are the expected *percentage* increase in the outcome variable when the explanatory variable increases by 1 unit and when all other variables are held constant at their mean and all dummy variables at 0.

Table 3.1 presents summary statistics for entire sample. In this sample, the average research article has garnered 71.4 citations until August 2017, has been written by 2.2 authors and consists of 26.9 pages. The authors' combined Euclidean Index equals 184.9. They were engaged in a total of 3.5 author-normalized projects (excluding the present project) and have a joint experience of 18.6 years. Negative values in total author experience are due to the construction of the variables, which are measured in the year before publication. If for example an article is the first for all authors, they have a negative experience. This is the case for 291 publications. Correlation coefficients are presented in table 3.B.2 (in the appendix to this chapter). We find a weak positive relationship between an article's citation count and author ability measured as sum of Euclidean indices. Correlation coefficients of total citations with the number of authors or the number of pages are equally weak. Most importantly, changes in the best commenter's centrality changes are not correlated with their number of projects or even the publication's citation count.

Table 3.2 reports marginal effects for all variables. Column (1) serves as reference model and excludes any commenter-related variables. All variables except the authors' combined experience and the authors' combined experience squared are statistically and economically significant. For example, each additional solo-authored project of the authors is associated with 4.9% more citations for the average paper. These coefficients do not change significantly across different specifications.

In column (2) we add characteristics of the most Bonacich central commenter: her total number of ongoing author-normalized research projects. The coefficient is statistically significant, as predicted by our model. For each additional ongoing single-authored project of the most central commenter from sample mean, citation count is expected to increase by 4.8%. This coefficient is about 30% higher than that for authors' combined author-normalized projects, partly accounting for the fact that the mean is lower.

Column (3) finally adds variable *Best. Com. Bonacich diff.*, which is the Bonacich centrality rank differential of the most Bonacich central commenter between the network of informal intellectual collaboration without and with deceased author. The coefficient of 61.50 is statistically significant with a p-value of 0.039. The coefficient is very high because of the very low mean of 0.00003 at which the coefficient is evaluated. The interpretation is as follows: For each increase in centrality by 1 from the mean, citations are expected to increase by 61.5 percent. To put this into perspective, if the average most central commenter (mean Bonacich centrality score: 0.0011) increases her centrality score by 1%, that is $0.01 * 0.0011 = 0.000011$, citations are expected to increase by $0.00011 * 61.6 \approx 0.6\%$ for the average article. This translates into $0.006 * 71.4 \approx 0.5$ more citations.

In column (4) we additionally control for the number of acknowledged seminars and conferences, as these might be correlated acknowledgment behavior. If no seminars or conferences are explicitly acknowledged, we assume this number to be 0. The coefficient of interest decreases somewhat to 85.9, but the effect size remains virtually constant.

Table 3.1: Summary statistics for all continuous variables.

	N	Mean	Median	Std.Dev.	Min	Max
Impact measure						
Total citations	4015	71.4	36	111.88	0	2163
Article characteristics						
# of pages	4015	26.9	27	10.42	3	128
# of authors	4015	2.2	2	0.84	1	6
Author characteristics						
Auth. total Euclid	4015	184.9	61	378.77	0	7590
Auth. total projects	4015	3.5	3	2.64	0	31
Auth. total experience	4015	18.6	15	16.13	−3	100
Auth. total experience^2	4015	604.9	225	956.83	0	10000
Best commenter characteristics						
Best Com. projects	4015	2.2	2	1.97	0	18
Best Com. experience	4015	18.5	15	77.42	−8	2010
Best Com. experience^2	4015	6335.0	225	155425.19	0	4040100
Network change						
Best. Com. Bonacich diff.	4015	0.0000326	0.0000225	0.000489	−0.0262	0.00101

Notes: # of authors and *# of pages* is the simple count of authors and pages, respectively. *Auth. total Euclid* is the author's Euclidean index as defined in (2.1) in the year before publication, summed over all authors. *Auth. total projects* is the number of other publications in the year of publication and the following year, divided by the number of co-authors, summed over all authors. *Auth. total experience* is the combined number of years between the year before publication and the publication year of each author's first article. *Auth. total experience*^{2} is its square. *Best Com. projects* is the number of publications in the year before the publication and the following year, divided by the number of co-authors, for the most central commenter. *Best. Com. experience* is the number of years between the year before publication and the publication year of the first article for the most central commenter. *Best Com. experience*^{2} is its square. *Best. Com. Bonacich diff.* is the difference of Bonacich centrality (equation (3.6)) measured in the giant component of the network of informal intellectual collaboration without and with deaths of the most central commenter, in the year before publication.

Table 3.2: Results of Negative Binomial regression for citation count.

	(1)	(2)	(3)	(4)
# of pages	0.013***	0.013***	0.013***	0.009***
	$p = 0.000$	$p = 0.000$	$p = 0.000$	$p = 0.0002$
# of authors	0.073***	0.074***	0.073***	0.107***
	$p = 0.002$	$p = 0.002$	$p = 0.002$	$p = 0.0005$
Auth. total Euclid	0.001***	0.001***	0.001***	0.001***
	$p = 0.000$	$p = 0.000$	$p = 0.000$	$p = 0.000$
Auth. total projects	0.050***	0.047***	0.047***	0.045***
	$p = 0.000$	$p = 0.000$	$p = 0.000$	$p = 0.00000$
Auth. total experience	−0.003	−0.002	−0.002	−0.001
	$p = 0.206$	$p = 0.333$	$p = 0.337$	$p = 0.798$
Auth. total experience^2	−0.0001	−0.0001*	−0.0001*	−0.0001**
	$p = 0.123$	$p = 0.069$	$p = 0.068$	$p = 0.012$
Best Com. projects		0.048***	0.048***	0.029***
		$p = 0.000$	$p = 0.000$	$p = 0.005$
Best Com. experience		−0.0003	−0.0004	0.003
		$p = 0.870$	$p = 0.807$	$p = 0.117$
Best Com. experience^2		0.00000	0.00000	−0.00000
		$p = 0.858$	$p = 0.795$	$p = 0.150$
Best. Com. Bonacich diff.			61.500**	58.900**
			$p = 0.039$	$p = 0.046$
# of seminars				0.028***
				$p = 0.000$
# of conferences				0.029***
				$p = 0.004$
Constant	3.690***	3.610***	3.610***	3.320***
	$p = 0.000$	$p = 0.000$	$p = 0.000$	$p = 0.000$
Publication year-fixed effects	*Yes*	*Yes*	*Yes*	*Yes*
Journal-fixed effects	*Yes*	*Yes*	*Yes*	*Yes*
N	4,015	4,015	4,015	2,171
Log Likelihood	-20,143.000	-20,124.000	-20,122.000	-11,120.000

Notes: ***, ** and * indicate statistical significance to the 1, 5 and 10 percent level, respectively. Reported coefficients are marginal effects. *# of authors* and *# of pages* is the simple count of authors and pages, respectively. *Auth. total Euclid* is the author's Euclidean index (2.1) in the year before publication, summed over all authors. *Auth. total projects* is the number of other publications in the year of publication and the following year, divided by the number of co-authors, summed over all authors. *Auth. total experience* is the combined number of years between the year before publication and the publication year of each author's first article. *Auth. total experience^2* is its square. *Best Com. projects* is the number of publications in the year before the publication and the following year, divided by the number of co-authors, for the most central commenter. *Best. Com. experience* is the number of years between the year before publication and the publication year of the first article for the most central commenter. *Best Com. experience^2* is its square. *Best. Com. Bonacich diff.* is the difference of Bonacich centrality (equation (3.6)) measured in the giant component of the network of informal intellectual collaboration without and with deaths of the most central commenter, in the year before publication.

3.6 Robustness Checks

3.6.1 Different specifications of Bonacich centrality

To test the robustness of our results, we alter the attenuation factor of the Bonacich centrality. We proceed to explore the importance of the attenuation factor α in equation (3.6). The attenuation factor governs the importance of distant links/nodes in the computation of the centrality of the focal node. Lower attenuation factor give less weights to distant nodes. When the attenuation factor reaches 0, only immediate neighbors are taken into account for the centrality computation.

In our main specification we have used an attenuation factor of attenuation factor of $0.85 \times 1/\mu_1(G_t)$, where $\mu_1(G_t)$ is the leading eigenvalue of the adjacency matrix of network G for year t (we scale with with eigenvalue to allow comparison between networks). In 3.3 we present regression results with the same specification as presented above in model 3.7. In model (1) we use an attenuation factor $0.99 \times 1/\mu_1(G_t)$, in model (2) an $0.90 \times 1/\mu_1(G_t)$, in model (3) the attenuation factor equals $0.80 \times 1/\mu_1(G_t)$, and in model (4) an attenuation factor of $0.70 \times 1/\mu_1(G_t)$.

Coefficients peak in magnitude at an attenuation factor of $0.8 \times 1/\mu_1(G_t)$ (model 3), while statistical significance is highest for an attenuation factor of $0.9 \times 1/\mu_1(G_t)$ (model 2). In our model, the attenuation factor is an exogenous variable and is interpreted as the contributive strength of complementarities in research efforts to individual utility. In equilibrium, it measures the decay in the contribution of equilibrium efforts of the first-order, second-order, third-order collaborators/neighbors, and so on, to equilibrium effort of the respective scientist. The smaller the attenuation factor the more the effort of first-order, second-order neighbors matter compared to that of distant neighbors. The value of the attenuation factor that gives the maximum and most significant effect of Bonacich centrality on citations then best describes the effects of complementarities on individual behavior. In our sample, this value is between $0.8 \times 1/\mu_1(G_t)$ and $0.9 \times 1/\mu_1(G_t)$.

Table 3.3: Results of Negative Binomial regression for citation count with different Bonacich centrality specifications.

	(1)	(2)	(3)	(4)
# of pages	0.013***	0.013***	0.013***	0.013***
	$p = 0.000$	$p = 0.000$	$p = 0.000$	$p = 0.000$
# of authors	0.071***	0.073***	0.073***	0.072***
	$p = 0.002$	$p = 0.002$	$p = 0.002$	$p = 0.002$
Auth. total Euclid	0.001***	0.001***	0.001***	0.001***
	$p = 0.000$	$p = 0.000$	$p = 0.000$	$p = 0.000$
Auth. total projects	0.046***	0.046***	0.047***	0.046***
	$p = 0.000$	$p = 0.000$	$p = 0.000$	$p = 0.000$
Auth. total experience	−0.002	−0.002	−0.002	−0.003
	$p = 0.407$	$p = 0.352$	$p = 0.339$	$p = 0.316$
Auth. total experience^2	−0.0001*	−0.0001*	−0.0001*	−0.0001*
	$p = 0.059$	$p = 0.067$	$p = 0.067$	$p = 0.079$
Attenuation factor 0.99				
Best Com. projects	0.046***			
	$p = 0.000$			
Best Com. experience	−0.004**			
	$p = 0.028$			
Best Com. experience^2	0.00000**			
	$p = 0.025$			
Centrality diff.	42.004*			
	$p = 0.087$			
Attenuation factor 0.90				
Best Com. projects		0.046***		
		$p = 0.000$		
Best Com. experience		−0.001		
		$p = 0.465$		
Best Com. experience^2		0.00000		
		$p = 0.459$		
Centrality diff.		63.330**		
		$p = 0.029$		
Attenuation factor 0.80				
Best Com. projects			0.046***	
			$p = 0.000$	
Best Com. experience			−0.0004	
			$p = 0.825$	
Best Com. experience^2			0.00000	
			$p = 0.814$	
Centrality diff.			63.868**	
			$p = 0.039$	
Attenuation factor 0.70				
Best Com. projects				0.047***
				$p = 0.000$
Best Com. experience				−0.001
				$p = 0.456$
Best Com. experience^2				0.00000
				$p = 0.448$
Centrality diff.				64.057*
				$p = 0.054$
Best Com. projects	3.667***	3.626***	3.611***	3.628***
	$p = 0.000$	$p = 0.000$	$p = 0.000$	$p = 0.000$
Publication year-fixed effects	*Yes*	*Yes*	*Yes*	*Yes*
Journal-fixed effects	*Yes*	*Yes*	*Yes*	*Yes*
N	4,015	4,015	4,015	4,015
Log Likelihood	-20,124.990	-20,124.000	-20,123.140	-20,123.020

Notes: ***, ** and * indicate statistical significance to the 1, 5 and 10 percent level, respectively. Reported coefficients are marginal effects. *# of authors* and *# of pages* is the simple count of authors and pages, respectively. *Auth. total Euclid* is the author's Euclidean index (2.1) in the year before publication, summed over all authors. *Auth. total projects* is the number of other publications in the year of publication and the following year, divided by the number of co-authors, summed over all authors. *Auth. total experience* is the combined number of years between the year before publication and the publication year of each author's first article. *Auth. total experience^2* is its square. *Best Com. projects* is the number of publications in the year before the publication and the following year, divided by the number of co-authors, for the most central commenter. *Best. Com. experience* is the number of years between the year before publication and the publication year of the first article for the most central commenter. *Best Com. experience^2* is its square. *Best. Com. Bonacich diff.* is the difference of Bonacich centrality (equation (3.6)) measured in the giant component of the network of informal intellectual collaboration without and with deaths of the most central commenter, in the year before publication.

3.6.2 Different network definition

Table 3.4 presents regression results similar to the baseline regression but with different centrality and network definitions.

A centrality measure that is closely related to Katz-Bonacich centrality is Eigenvector centrality. When the starting vector $\mathbf{t}$ is set to 0, so that there are no initial centralities, and the attenuation factor is set equal to $1/\mu_1(A)$, an equivalence exists between Bonacich and Eigenvector centralities. Model 1 of table 3.4 presents results for our main specification with Eigenvector centrality instead of Bonacich centrality. The coefficient of 41.600 is much lower, while the corresponding p value of 0.091 is nearly beyond conventional significance thresholds. We interpret this result as implying that the significance of weighted Bonacich centrality $\mathbf{b}(A,\alpha,\mathbf{t})$ in our main results matter.

Our model predicts that spillovers depend on frequency of interaction and types of the participating scholars. In column (2) of table 3.4 we compute Bonacich centrality in networks without any link weights. The coefficient becomes turns statistically insignificant. We take this as sign that accounting for the frequency of interactions among scientists and the types of the participating scientists indeed matters in capturing strategic complementarities.

Our results show that connections in the network of informal intellectual collaboration matter. In column (3) of table 3.4 we compute and compare Bonacich centrality in networks of informal and formal collaboration. That is, two academics are connected not only when one acknowledges the other but also when they have jointly published a paper. In this case the link weight increases by one for each joint publication, and as before by $1/n$ for each acknowledgement on a publication with n authors. The coefficient of comparable size as compared to main results. This is due to the lower sample mean of the variable, at which the coefficient is evaluated in a negative binomial regression. The mean change in Bonacich centrality of the best commenter in networks of formal and informal collaboration is 0.000023, which is about 30% lower than the mean change of Bonacich centrality in the network of informal collaboration.

Table 3.4: Results of Negative Binomial regression for citation count with different network definitions.

	(1)	(2)	(3)
# of pages	0.013^{***}	0.013^{***}	0.013^{***}
	$p = 0.000$	$p = 0.000$	$p = 0.000$
# of authors	0.072^{***}	0.075^{***}	0.073^{***}
	$p = 0.002$	$p = 0.001$	$p = 0.002$
Auth. total Euclid	0.001^{***}	0.001^{***}	0.001^{***}
	$p = 0.000$	$p = 0.000$	$p = 0.000$
Auth. total projects	0.046^{***}	0.048^{***}	0.046^{***}
	$p = 0.000$	$p = 0.000$	$p = 0.000$
Auth. total experience	-0.002	-0.003	-0.002
	$p = 0.430$	$p = 0.187$	$p = 0.335$
Auth. total experience^2	-0.0001^{*}	-0.0001	-0.0001^{*}
	$p = 0.053$	$p = 0.157$	$p = 0.068$
Eigenvector centrality			
Best Com. projects	0.050^{***}		
	$p = 0.000$		
Best Com. experience	-0.004^{**}		
	$p = 0.027$		
Best Com. experience^2	0.00000^{**}		
	$p = 0.027$		
Centrality diff.	41.600^{*}		
	$p = 0.091$		
Unweighted network			
Best Com. projects		0.027^{***}	
		$p = 0.0002$	
Best Com. experience		-0.006^{***}	
		$p = 0.0002$	
Best Com. experience^2		0.00000^{***}	
		$p = 0.0003$	
Centrality diff.		-36.900	
		$p = 0.145$	
Network with formal collaboration			
Best Com. projects			0.044^{***}
			$p = 0.000$
Best Com. experience			-0.0002
			$p = 0.907$
Best Com. experience^2			0.00000
			$p = 0.902$
Centrality diff.			68.100^{**}
			$p = 0.020$
Constant	3.660^{***}	3.740^{***}	3.610^{***}
	$p = 0.000$	$p = 0.000$	$p = 0.000$
Publication year-fixed effects	*Yes*	*Yes*	*Yes*
Journal-fixed effects	*Yes*	*Yes*	*Yes*
N	4,015	4,015	4,015
Log Likelihood	-20,123.000	-20,130.000	-20,123.000

Notes: ***, ** and * indicate statistical significance to the 1, 5 and 10 percent level, respectively. Reported coefficients are marginal effects. *# of authors* and *# of pages* is the simple count of authors and pages, respectively. *Auth. total Euclid* is the author's Euclidean index (2.1) in the year before publication, summed over all authors. *Auth. total projects* is the number of other publications in the year of publication and the following year, divided by the number of co-authors, summed over all authors. *Auth. total experience* is the combined number of years between the year before publication and the publication year of each author's first article. *Auth. total experience^2* is its square. *Best Com. projects* is the number of publications in the year before the publication and the following year, divided by the number of co-authors, for the most central commenter. *Best. Com. experience* is the number of years between the year before publication and the publication year of the first article for the most central commenter. *Best Com. experience^2* is its square. *Best. Com. Bonacich diff.* is the difference of Bonacich centrality (equation (3.6)) measured in the giant component of the network of informal intellectual collaboration without and with deaths of the most central commenter, in the year before publication.

3.6.3 Betweenness centrality

The equilibrium behavior of our model shows that Bonacich centrality best captures individual influence in an environment of intellectual collaboration with positive complementarities. The existence of complementarities is necessary for Bonacich centrality to be an appropriate measure of influence in equilibrium. If the underlying process driving our empirical results were pure information flow, then other centrality measures can equally capture a scientist's level of influence. In particular, the betweenness centrality (equation (2.5)) is so often used to measure the level of individual influence in relation to information flows within a network Freeman (1978).

Although both measures captures an agent's influence in terms of information possessed in the process of information flow, Bonacich centrality is unique to the process of strategic interactions with positive complementarities. Hence, if scientists' betweenness centralities significantly influence their research output, then our results above could also be a result of pure information contagion and not existence of positive complementarities.

In Table 3.5 we test whether betweenness centrality of commenters is statistically significantly associated with an article's citation count.

The explanatory variable is the difference in betweenness centrality of the most betweenness central commenter between the network of informal intellectual collaboration without and with authors that deceased during the previous periods. Though the coefficient is high, it is not statistically significant. This indicates that the positive and significant result we obtain for Bonacich centrality is at the very least partially driven by strategic complementarities and not just pure information spillovers.

Table 3.5: Results of Negative Binomial regression for citation count with betweenness centrality.

	(1)
# of pages	0.012^{***} $p = 0.000$
# of authors	0.076^{***} $p = 0.001$
Auth. total Euclid	0.001^{***} $p = 0.000$
Auth. total projects	0.047^{***} $p = 0.000$
Auth. total experience	-0.002 $p = 0.360$
Auth. total experience^2	-0.0001^{*} $p = 0.063$
Best Com. projects	0.047^{***} $p = 0.000$
Best Com. experience	-0.003^{*} $p = 0.091$
Best Com. experience^2	0.00000^{*} $p = 0.082$
Betweenness centrality diff.	55.200 $p = 0.392$
Constant	3.660^{***} $p = 0.000$
Publication year-fixed effects	Yes
Journal-fixed effects	Yes
N	4,015
Log Likelihood	-20,123.000

Notes: ***, ** and * indicate statistical significance to the 1, 5 and 10 percent level, respectively. Reported coefficients are marginal effects. *# of authors* and *# of pages* is the simple count of authors and pages, respectively. *Auth. total Euclid* is the author's Euclidean index (2.1) in the year before publication, summed over all authors. *Auth. total projects* is the number of other publications in the year of publication and the following year, divided by the number of co-authors, summed over all authors. *Auth. total experience* is the combined number of years between the year before publication and the publication year of each author's first article. *Auth. total experience*^{2} is its square. *Best Com. projects* is the number of publications in the year before the publication and the following year, divided by the number of co-authors, for the most betweenness central commenter. *Best. Com. experience* is the number of years between the year before publication and the publication year of the first article for the most betweenness central commenter. *Best Com. experience*^{2} is its square. *Best. Com. betweenness diff.* is the difference of betweenness centrality (equation (2.5)) measured in the giant component of the network of informal intellectual collaboration without and with deaths of the most betweenness central commenter, in the year before publication.

3.6.4 Do commenters cite the paper more?

An important concern is whether a higher number of acknowledged commenters increases citation count because of dissemination of the publication. That is, authors advertise the manuscript while informally collaborating. Speaking to more researchers hence broadens the audience and thus more informal collaboration translates into higher citation counts.

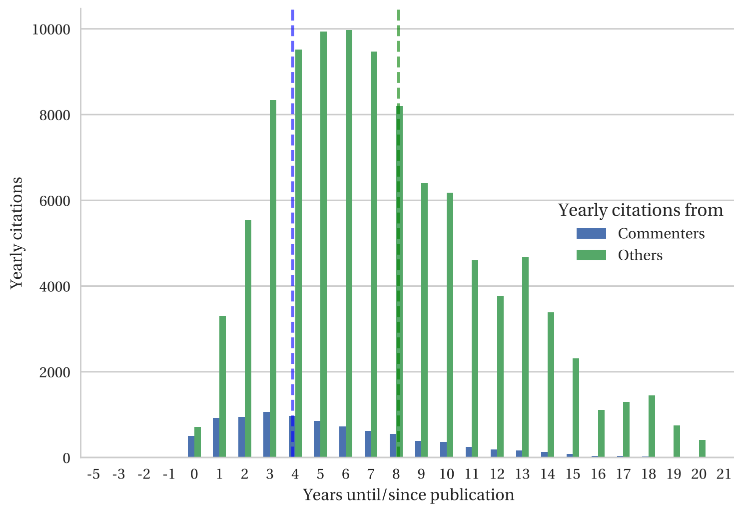
We assess the importance of this channel. For each article in our dataset, we look up whether citing authors are also acknowledged on the article they cite (this information orginates from Scopus). The total number of citations of the 4105 articles in our dataset equals $492,636$. 1.8% of the citations are due to articles published by acknowledged commenters (the number is slightly lower when including joint publications by acknowledged commenters and original author). Out of $4,105$ articles, $2,905$ are never cited by its commenters, 852 articles are cited once by a commenter, 236 articles are cited twice by a commenter, 79 articles are cited thrice, and 33 articles are cited four or more times by its acknowledged commenters. Given that an article is cited by its commenters, the average article is cited by 24% of its commenters (unconditional: 11%).

These commenters however do tend to cite the article shortly after the publication, as 3.2 reveals. Half of the citations from commenters' publications occur after 4 years of the publication, while 50% of the citations from non-commenters' publications occur after 8 years. The weighted average time until a citation from a commenter occurs is 5.23 years, while that for Others is 8.39.[8]

Only a small portion of acknowledged commenters cite the article they are acknowledged on. Figure 3.3 shows on the y-axis that on average 11% of all acknowledged commenters on an article eventually cite it (where the citing article must be without the original authors). As the total number of acknowledged commenters varies, the x-axis gives an account of the absolute number of acknowledged citing commenters. Its mean is 1.

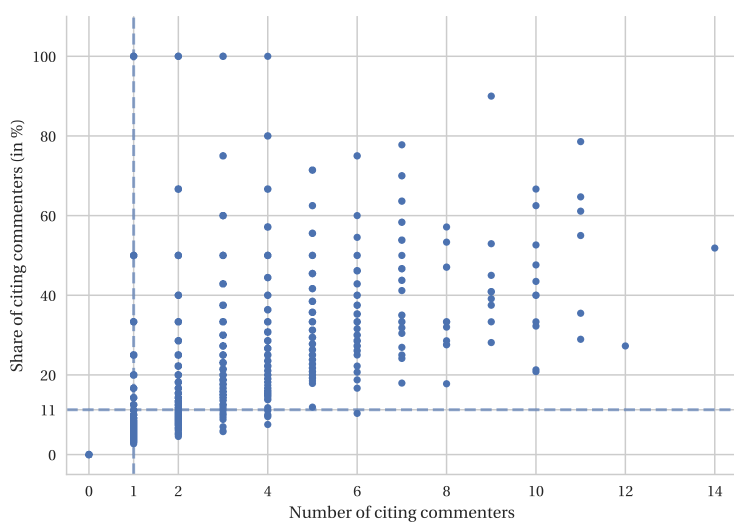
[8]The measure corresponds to Macaulay duration used to characterize bonds. We define it as $\sum_t^n t * c_t / \sum_t^n c_t$ for n periods before and after publication year in which the publication received at least one citation. c_t is the number of citations in period t.

Figure 3.2: Barplot showing when papers are cited by its acknowledged commenters, by number of how many commenters cite it.



Notes: Figure indicates the time lag before/since publication commenters cite the publication they are acknowledged on, as compared to non-commenters that cite the article. Dashed lines indicate the lag until which 50% of the citations from that group occur. The total number of papers is 5320. A citation from "Commenters" means that at least one of the authors is acknowledged commenter and none of the authors is authoring the cited publication. A citation from "Others" includes self-citations.

Figure 3.3: Scatterplot showing the amount of commenters that cite the publication.



Notes: Figure compares the number of acknowledged commenters that cite the article they are acknowledged on (citing commenter), versus its share. Dashed lines indicate the means. To count as citing commenter, the commenter must coauthor a publication where she is acknowledged on as commenter, but without the original authors.

Based on these numbers we conclude that the citations from acknowledged commenters play a minor role, and that our measure of impact is not confounded by dissemination effects. Additional channels, that we do not consider here, are dissemination through conferences and seminars. This channel is likely to increase the share of citations from informal collaborators. On the other hand, it is unclear whether informal collaborators would have cited the article anyways had they not seen the article before.

3.7 Conclusion

Researchers that collaborate either formally or informally inevitably diffuse information for example on new ideas, emerging trends and upcoming challenges. The extent to which individuals participate in the diffusion process depends on their position in the social network of intellectual collaboration (Jackson, 2014). We show that research articles benefit more from a commenter's comments when the commenter is more Bonacich central, i.e. when she is closer to the most connected clique in the network (Bonacich, 1987). An increase by 2% in Bonacich centrality of the average most central commenter on a research article increases citation count by ~ 1 citations for the average article.

Overall, our stylized model and empirical analysis highlight the importance of intellectual collaboration and network effects on the impact academic work can have. The importance of intellectual collaboration adds new insights into the division of labor in academic teams. There is a wide range of activities that are necessary for scientific innovation (Haeussler and Sauermann, 2016). But not all of these need to be performed by co-authors only, i.e. authors in economics can extend the team to outsource activities that do not justify co-authorship. For example, authors test arguments and the scope of their article's contribution while presenting, or they rely on trusted assessors for relevant literature. It is precisely these larger groups that we target at.

Appendix

3.A Proof of Proposition 1

Each scientist chooses a level of effort that maximizes (3.2). The respective first order condition is

$$\beta e_i^* = t_i + \alpha \sum_{j \in N_i} g_{ij} e_j^* \quad \text{for each } i \in N. \tag{3.8}$$

Writing $\mathbf{e}^*$ for the row-vector of equilibrium efforts, and recalling that G is the weighted adjacency matrix, then (3.8) can be written in matrix form as

$$\beta \mathbf{e}^* - \alpha \mathbf{e}^* G = \mathbf{e}^* \left(\beta I - \alpha G\right) = \mathbf{t}$$

Debreu and Herstein (1953, Theorems III^* and III) show that the matrix $\left(\beta I - G\right)$ is well-defined and non-negative, that is $\left(\beta I - G\right) > 0$, whenever $\beta > \mu_1(G)$. They also show that under such conditions, for any pair of vectors $\mathbf{x}$ and $\mathbf{y} \geq 0$, such that $\mathbf{x} = \left(\beta I - G\right)^{-1} \mathbf{y}$, then $\mathbf{x} > 0$. This then implies that

$$\mathbf{e}^* = \frac{1}{\beta} \left(I - \frac{\alpha}{\beta} G\right)^{-1} \mathbf{t} = \frac{1}{\beta} \mathbf{b}(G, \alpha_\beta, \mathbf{t}) \tag{3.9}$$

is an interior equilibrium vector whenever $\beta > \alpha \mu_1(G)$. The proof of uniques, that is the non-existence of corner solution, is identical to the proof of Theorem 1 Ballester et al. (2006).

3.B Additional tables and figures

Table 3.B.1: List of deceased scholars used for identification in chapter 2.

	Date of death		Date of death
Aiyagari, S. Rao	1997, May 20	Fini, Riccardo	2007, January 20
Bailey, Martin J.	1997, June 26	McMillan, John	2007, March 13
Maddala, G. S.	1999, June 04	Sokoloff, Kenneth L.	2007, May 21
Marrinan, Jane	2000, January 02	Agell, Jonas	2007, July 01
Rosen, Sherwin	2001, March 17	Barclay, Michael J.	2007, August 16
West, Edwin G.	2001, October 06	Glyn, Andrew J.	2007, December 22
Tobin, James	2002, March 11	Benston, George J.	2008, February 01
Smith, Bruce D.	2002, July 09	Cass, David	2008, April 15
Dornbusch, Rüdiger	2002, July 25	Bartolini, Leonardo	2008, July 09
Gabriel, Stuart A.	2002, October 15	Terrell, Katherine	2009, January 01
Kindleberger, Charles P.	2003, July 07	Neftçi, Salih N.	2009, April 15
Flemming, John S.	2003, August 05	Granger, Clive W.J.	2009, May 27
Spulber, Nicolas	2004, January 02	Prati, Alessandro	2009, June 21
Lee, Winson	2004, March 01	Weston, J. Fred	2009, August 01
Laffont, Jean Jacques	2004, May 01	Mozumdar, Abon	2009, November 05
Freeman, Scott D.	2004, July 23	Samuelson, Paul A.	2009, December 13
Berkowitz, Michael K.	2004, August 08	Auernheimer, Leonardo	2010, January 01
Grossman, Herschel I.	2004, October 09	Ghosh, Dipak	2010, January 10
Battalio, Raymond C.	2004, December 01	Stockman, Alan C.	2010, January 14
Bergstrom, Albert Rex	2005, May 01	Dickhaut, John W.	2010, April 10
Hirshleifer, Jack	2005, July 26	Shastri, Kuldeep	2010, April 19
Xia, Yihong	2005, August 06	Hirschey, Mark John	2010, July 18
Geroski, Paul A.	2005, August 28	Zellner, Arnold	2010, August 11
Branson, William H.	2006, August 15	McKenzie, Lionel W.	2010, October 12
Billett, Matthew T.	2006, September 14	Kim, Bonghan	2011, January 01
Xue, Hui	2006, September 14	Urban, Dieter M.	2011, March 07
Fischer, Klaus P.	2006, October 06	Schneller, Meir I.	2011, April 10
Saxonhouse, Gary R.	2006, November 01	Howrey, E. Philip	2011, June 17
Friedman, Milton	2006, November 16	Warga, Arthur D.	2011, August 07
Goldberg, Lawrence G.	2007, January 01	Orbay, Hakan	2011, September 15
Kandel, Shmuel	2007, January 01		

Notes: Table lists names and dates of death for all researchers that occur as author or acknowledged commenter in our networks.

Table 3.B.2: Spearman and Pearson correlations for all continuous variables.

Impact measure									
Total citations		1.00	0.36	0.20	0.08	1.00	0.07	0.56	0.16
Article characteristics									
# of pages	0.24		0.13	0.12	−0.05	0.05	−0.02	0.47	0.56
# of authors	0.07	0.05		0.05	0.11	0.16	0.14	0.62	1.00
Author characteristics									
Auth. total Euclid	0.16	0.10	0.28		0.36	0.01	0.13	0.00	0.51
Auth. total projects	0.07	−0.01	0.39	0.36		0.03	0.05	0.02	0.82
Auth. total experience	0.03	0.02	0.60	0.52	0.41		1.00	0.09	0.04
Commenter characteristics									
Best Com. projects	0.06	0.04	0.00	0.03	0.08	−0.01		0.20	0.07
Best Com. experience	−0.01	−0.02	0.04	0.00	−0.02	0.01	−0.01		0.14
Network measure									
Best. Com. Bonacich diff.	0.02	0.00	0.02	0.04	0.01	0.02	0.04	0.00	

Notes: Upper triangular shows Spearman (rank) correlation coefficients, while lower triangular shows Pearson correlation coefficients. *# of authors* and *# of pages* is the simple count of authors and pages, respectively. *Auth. total Euclid* is the author's Euclidean index (2.1) in the year before publication, summed over all authors. *Auth. total projects* is the number of other publications in the year of publication and the following year, divided by the number of co-authors, summed over all authors. *Auth. total experience* is the combined number of years between the year before publication and the publication year of each author's first article. *Auth. total experience*^{2} is its square. *Best Com. projects* is the number of publications in the year before the publication and the following year, divided by the number of co-authors, for the most central commenter. *Best. Com. experience* is the number of years between the year before publication and the publication year of the first article for the most central commenter. *Best Com. experience*^{2} is its square. *Best. Com. Bonacich diff.* is the difference of Bonacich centrality (equation (3.6)) measured in the giant component of the network of informal intellectual collaboration without and with deaths of the most central commenter, in the year before publication.

Chapter 4

Informal Contacts in Hiring: The Economics Job Market

4.1 Introduction

The Economics labor market is organized annually by the American Economic Association with graduate students, their advisers and hiring institutions as market participants (Coles et al., 2010). Matching students to universities remains imperfect due to, inter alia, asymmetric information about the quality of the applicant. Reference letters and phone calls from the adviser(s) reduce these asymmetries (Athey et al., 2016; Colander, 1997).[1] Since an adviser can reduce the information asymmetry about the quality of her students in her academic network[2], we expect that better 'connected' advisers are able to reduce this information asymmetry more than less connected advisers. We test the hypothesis that PhD students of well connected advisers obtain better first placements than those of less connected advisers in the academic market for Economics graduates. We define adviser connectedness as the Eigenvector centrality rank of

[1] For example, Colander (1997) writes that "Recommendations from important people are extremely important" and "Informal contacts - and phone calls by your advisers and friends - are important".

[2] One other example of using social networks to reduce information asymmetry is given by Baruffaldi et al. (2016), who show that PhD students who obtained their Master's degree at an affiliation of their adviser's co-author are more productive than PhD students coming from a university to which their adviser has no links.

the adviser in the collaboration frequency weighted network of co-authors in the field of Economics, and we create our own ranking based on the methodology of the Tilburg University Economics ranking to judge the quality of a student's placement university. Eigenvector centrality measures not just 'connectedness' in terms of direct connections (number of co-authors), but also allows for indirect connections to influence connectedness. For example, if an adviser's co-author becomes more connected then this adds to the connectedness of the adviser as well, and this could potentially influence the placement of her students.[3]

The contribution of this chapter can be separated into three parts. Oyer (2006) points out that the first placement of graduate students has a significant impact on their careers. Thus, first and foremost, our research is important for graduate students since it demonstrates another channel through which an adviser can influence student placement. The role of the adviser has also been explored by Krueger and Wu (2000), who report a correlation between the subjective prominence of the letter writer and student placement. We are also able to confirm that 'prominence' (which we in contrast equate with the Euclidean index[4] of citations) matters, but it does not account completely for the impact of adviser connectedness. We show that even after controlling for prominence, the connectedness of the adviser matters for her students' placement.

Secondly, though our results are for a unique job market, we are confident that our results offer two relevant insights for the general labor market as well. One, owing to our unique data set we are able to demonstrate the importance of even indirect connections in job search. For example, though an adviser may not have a co-author at a given University, she may still be able to put in a good word for her student there if her coauthor has a co-author in that university. This distinguishes our chapter from those in the literature on referrals who only look at direct

[3]Eigenvector centrality is the weighted sum of the Eigenvector centralities of the immediate neighbors, where the weights correspond to the neighbors' Eigenvector centralities. The idea behind this measure is that people connected to more connected individuals are themselves more connected. This has been shown to be informative in various settings. Cruz et al. (2017) recently show that politicians who are more Eigenvector central in a network of families receive higher voter turnout. Similarly, Calvó-Armengol et al. (2009) show that equilibrium efforts in networks are proportional to a variant of the Eigenvector centrality, and Banerjee et al. (2013) show that the Eigenvector centrality of the first-informed individual predicts how fast information spreads in a social network. A further discussion of this measure is in section 4.3.3.

[4]The Euclidean index of citations is the Euclidean sum of publications represented by their citation stock for any given year, as proposed by Perry and Reny (2016).

links (Kramarz and Skans, 2014; Burks et al., 2017) - current employees referring a worker in their current workplace. While these papers establish the value of direct connections, they do not speak about which workers are more 'connected' than others, and whether this matters for job seekers. For example, a link with a worker who has worked in several firms may be more valuable to a job seeker than a link with a worker who has worked at only one firm because the former has connections with more employers[5]. It is important to understand how connectedness impacts job outcomes because this could have repercussions on the distribution of income. People close to more connected workers may have higher incomes than people with fewer connections, and this could perpetuate. Two, by focusing on the Economics job market where there is little information asymmetry about new job openings (thanks to Job Openings for Economists and other web pages), we provide some supportive evidence to show that network connections can be used to reduce information asymmetry about the quality of the job candidate, and this helps applicants get better jobs. This is an important distinction from the literature on job search (Granovetter, 1973; Bayer et al., 2008), which is usually unable to distinguish between two channels via which social networks usually affect labor market outcomes: learning about new job openings versus reducing information asymmetry about the job candidate.[6]

A final contribution of this chapter is the novel data set we create for our analysis, which was collected from various first and second hand sources. Our sample consists of 3,182 Economics students who obtained their PhD from 137 different North American universities during the academic years 2000/2001, 2001/2002, 2002/2003 and 2003/2004. Our data originates from the Journal of Economic Literature, which in its December issues publishes a list of new Economics graduates of North American universities. For each student, we find the adviser and the

[5] Assuming of course that the worker left his past employers on good terms! Also, while we realize that the general labor market does not have an 'adviser' who places workers, at its heart, this chapter is about pointing out that more connected people may be more valuable in a job search. This message will be true for any labor market with information frictions.

[6] There are, however, theoretical arguments highlighting the result that referrals by current employees can diminish information problems arising from the fact that employers do not know worker quality perfectly (Montgomery, 1991; Burks et al., 2017; Dustmann et al., 2016; Hensvik and Skans, 2016). Furthermore, while economics graduates know about most of the job openings, they may not have other relevant information regarding these openings such as the work environment at the prospective department. An adviser can help reduce these information asymmetries as well. Thus, our focus on channel side is a reduction in information asymmetry - this could be regarding student quality or about other variables which can affect the match quality.

first placement from various sources including department websites, direct emails to departments, students' CVs, the genealogy project of the RePEc database and The Academic Family Tree project. The social network of coauthors is constructed using 114,409 publications in 408 journals.

We show that advisers who are more connected (have a better eigenvector centrality rank in the network of coauthors) place their students at better ranked universities compared to those who are not as well connected. Obviously, the connectedness of the adviser in the network of co-authors is endogenous. We identify the impact of adviser connectedness by using the *changes* in the connectedness of the adviser's co-authors in the year of student placement (in a model with adviser-fixed effects) as an instrument. The adviser's co-authors' connectedness is computed in the co-author network which excludes the adviser. This is to avoid changes in the adviser's own connectedness changing the connectedness of her coauthors. The identifying condition here rests on the following ideas. One, we control for time-invariant unobserved adviser characteristics via adviser-fixed effects. Two, *changes* in the connectedness of the adviser's co-authors *in the year of a student's placement* would be difficult to anticipate (and therefore use strategically) for both the student and the adviser. Therefore, it can be thought of as an exogenous variable which changes adviser connectedness in the year of placement, and affects student placement only via this channel. A deeper discussion of the critical empirical challenges and our identification strategy is presented in section 4.4.

In section 4.5.2, we provide additional evidence to support our hypothesis that an adviser's connectedness matters for her student's placement. We use the death of economists as an exogenous shock which affects the 'social distance'[7] between an adviser and a department, and we show that an increase in this social distance negatively affects the probability of the adviser placing her student at that department. Finally, in section 4.6, we provide supportive evidence to argue that the channel through which an adviser's connectedness affects her student's placement is that it helps the hiring university screen better by reducing information asymmetry regarding the student's quality.

[7] The shortest path between the adviser and the department in the co-author network.

4.2 Literature

Social networks and informal connections help a worker in finding a job in two main ways. One, by giving the worker information about new job postings (Granovetter, 1973; Boorman, 1975; Calvó-Armengol, 2004; Calvó-Armengol and Zenou, 2005). Two, by reducing the information asymmetry between the worker and the employer about the worker's ability[8] (Montgomery, 1991; Burks et al., 2017; Dustmann et al., 2016). However, studies which look at the former channel - like those which study the impact of neighborhoods and other 'local' networks on job search, often cannot distinguish between these two effects. For example, Bayer et al. (2008) find that individuals residing on the same block are more likely to work together. However, it is not clear if this is simply because neighbors learn from each other about new job openings, or if they actually recommended them. In contrast, we study the Economic job market where the job seekers have almost full information about all the job openings. This is because most job openings are posted on one web page - Job Openings for Economists (JOE). Thus, the setting in this chapter is particularly conducive to studying how much social connections can help by reducing information asymmetry about the quality of the job candidate.

The literature on referrals studies how firms can screen better when they use referrals from their current workers to hire new employees. Montgomery (1991) establishes that there are gains from referral hiring as employers can utilize recommendations from their productive workers to identify other productive potential workers. Hensvik and Skans (2016) directly test this empirically. Building on learning models, Dustmann et al. (2016) hypothesize that job search networks help reduce information deficiencies in the market and consequently referral-based job searches lead to better matches. The authors proceed to test the prediction empirically and show that referred workers initially earn higher wages. Few other studies have addressed this issue empirically. Burks et al. (2017) for example, show that referred workers are less likely to quit even though their productivity does not differ from that of non-referred hirings.

The literature on referrals makes clear that knowing a currently employed worker can in-

[8] Nepotism and reciprocity are other possible channels.

crease the probability of landing a job at their current firm for a job seeker. However, the papers in the literature do not quantify the value of any given connection as compared to another. For example, if a job seeker has a link with two workers, which one is more valuable to the worker? Do they have the same value if the two workers work in similar firms in similar positions? The chapter contributes to the literature on job search by using a unique data set to map the network of connections to show a causal relationship between nodes which are more connected (as per network theory measures of connectedness) and job market outcomes of job seekers who have a link with these nodes. Furthermore, since we study the entire network of social connections, we are able to take into account the impact of indirect connections as well as those of direct links. For example, Kramarz and Skans (2014) look at the direct link between parent and children and establish that this connection is important for the job market outcome of the children. Burks et al. (2017), Hensvik and Skans (2016) show that workers recommended by the current employees are often a better match for the firm. However, if a worker knows about a job opening at her previous employer or if a worker knows a friend who is employed in a different firm, then these papers will not be able to capture how the worker can use this information to help a job seeker get the job. For this, one would have to know the entire network of work-links amongst all workers. Some studies do try to infer networks from available data. For example, Dustmann et al. (2016) proxy a referral hire by the share of workers in the firm with the same ethnicity as the applicant. However, these inferences are imprecise. In contrast, our data allows us to precisely link economists whenever they have published a research article.

The literature on the Economics job market is relatively small. We know of no other papers which study the importance of the connectedness of the adviser for the placement of the student. In addition, most papers establish interesting correlations without showing causality. For example, Athey et al. (2016) look at graduates from the top PhD programs in the USA to show that first year (graduate school) grades in core courses of Microeconomics and Macroeconomics are significantly related to better job placement. They also report that the quality of the undergraduate institution of the student also affects the quality of first job. Krueger and Wu (2000) show that the 'prominence' (measured in an admittedly subjective manner) of the recommendation letter writer helps student placement. Smeets et al. (2006) show that market does not rely completely on the reputation of the PhD granting university since they find that often the top graduates of very good (but not elite) programs outperform average graduates of elite programs in terms of initial placement. Gallet et al. (2005) test a multi-market strategy in the Economics Job Market and confirm its predictions regarding cyclicity: “in the bust market, graduates of elite schools shifted their search strategies to include weaker academic institutions, while graduates of lower-ranked schools shifted their applications away from academia and toward the business sector." To avoid this kind of selection bias was one of the reasons why we chose to work with data from the years before the financial crisis for our study. Baruffaldi et al. (2016) study the use of academic networks (science and engineering students only) in the hiring of PhD students and its impact on student productivity. They show (without claiming causality) that PhD students hired from masters programs at affiliations from which the adviser of the PhD student draws co-authors have, on average, a higher productivity compared to students hired from universities to which their adviser has no links. Unlike our study though, this study is unable to indicate the importance of the centrality of the student’s masters professors on their PhD program placement.

4.3 Data

4.3.1 Doctoral Dissertations in Economics

Lists of doctoral students receiving their PhD from an Economics-related faculty in the US and Canada are published annually in the December issue by The Journal of Economic Literature (JEL). These dissertations are, with few corrections and additions as well as different information, also available from EconLit. We focus on four academic years, namely 2000/2001, 2001/2002, 2002/2003, and 2003/2004. The information from both the JEL lists and EconList include the JEL field of their dissertation along with the year in which they were awarded the PhD and the name of the PhD school.

We have information on 3,482 students from 137 different schools. To obtain a more homogeneous set, we remove 300 students that belong to JEL general category "Q" (Agricultural and Natural Resource Economics; Environmental and Ecological Economics) because the labor market for agricultural Economists is different from the rest of Economics. After this process we are left with 3,182 students, whose distribution of year and JEL code is summarized in table 4.1. The largest field is O (Economic Development, Innovation, Technological Change, and Growth) with 420 students. Four fields have more than 300 students: F (International Economics), G (Financial Economics), D (Microeconomics) and E (Macroeconomics and Monetary Economics).

Table 4.1: Crosstable by year and JEL code for all PhD students.

JEL Year	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	R	Z	All
2000	0	1	20	34	40	49	34	16	30	39	2	33	5	5	71	9	12	0	400
2001	0	2	50	76	82	99	95	38	55	78	7	67	11	11	94	18	18	0	801
2002	0	3	39	82	79	88	96	22	53	71	4	63	11	10	112	13	22	0	768
2003	1	3	39	93	89	85	79	34	68	63	6	54	9	11	95	13	17	1	760
2004	1	1	31	53	39	64	56	15	37	35	7	33	9	4	48	14	5	1	453
All	2	10	179	338	329	385	360	125	243	286	26	250	45	41	420	67	74	2	3182

Notes: Table lists numbers of graduated PhD students from North American universities for the academic years 2000/2001, 2001/2002, 2002/2003, and 2003/2004 by Journal of Economic Literature general category. Students from JEL general category Q ("Agriculture") are excluded.

We obtain information about the student's advisers from four sources. First, we use the genealogy database of the Research Papers in Economics (RePEc) project.[9] Second, we obtain adviser information from academic departments, either through public sources in the form of websites, or privately through direct emails.[10] Third, we collect CVs of the students themselves. The fourth source includes various online sources such as academic tree or Mathematics Genealogy Project. Using the Scopus database we compute the Euclidean index of citations for each year for each adviser as a measure of adviser productivity. Perry and Reny (2016) show that, unlike other indices (such as the h-index), this index has desirable properties if one is interested in combining citation stock and publication count.

Table 4.A.1 in the appendix ranks advisers by number of graduating students in 2000-

[9] See https://genealogy.repec.org/ Information on advisers requires the existence of a RePEc account of the student.

[10] Of 131 contacted departments, 29 sent information, 17 declined to share these information and 10 do not have records from the period 1999-2004.

2004 period. The ranking is lead by Daron Acemoglu having a total of 23 PhD students (which includes co-supervised students). He is followed by Andrej Shleifer (20 students) and Roger Betancourt (18 students). A data issue we face is that a fraction of the advisers in our sample had only one of their students graduate in the 2000/2001-2003/2004 period. Figure 4.A.1 in the appendix shows that out of 1,384 advisers, more than 650 advisers had only 1 student graduate in our time period, about 300 advisers had 2 students, and less than 180 advisers had 3 students graduating in this period. This implies that an analysis with adviser-fixed effects would have to use data from a subsample of students only. This will create a sample selection bias. Since we lose more students from lower ranked universities to this selection, and the impact of adviser connectedness should be higher for lower ranked universities,[11] we advise that our estimated impact of adviser connectedness on student placement be thought of as a conservative estimate. To use more data, we also report results from regressions without adviser-fixed effects. However, the identification condition needed in this regression is much stronger.

Finally, we estimate the gender of students and advisers based on their first name using the genderize.io database.[12] Out of the 1,384 advisers, we estimate 159 to be female, which corresponds to a share of roughly 10%. The share of female PhD students is higher, with 1,212 out of 3,182 (28%).

4.3.2 Economics Job Market

Information on initial placements is available either through the student's CV or from their former departments directly.[13] Figure 4.A.3 in the appendix visualizes the hiring network for the students in our network. To measure the quality of the initial placement we convert the initial placement into placement ranks. We use the *method* of the Tilburg Economics University Ranking to rank universities according to their research output. The Tilburg Economics Uni-

[11] This is because students from better ranked universities usually reduce information asymmetry about their quality via publications before the job market, and presenting their work at top conferences (this also gives them a chance to network).

[12] See `https://genderize.io/`.

[13] Not all universities give an account of their student's initial placement. We contacted all departments to share information. 31 sent placement information, 17 declined to share these information and 14 do not have records of that time. The remainder did not answer.

versity Ranking uses a weighted publication output in 74 predefined journals to assign points to the authors' main affiliation, and ranks universities according to these points[14]. Using data from Scopus and not the Web of Science, our version of the ranking covers more universities and colleges and thus allows a greater sample size. We use the Scimago Journal Impact Factor, rather than the Web of Science Journal Citation reports, in the year of the publication as journal weight because it was computed using data from Scopus and thus complements our data source.[15] Using data from Scopus allows us to include rank affiliations by organization type. In the main version of our ranking, we rank all affiliations which are classified as university or college. We use the affiliation rank in the year of the placement. Figure 4.A.2 (in the appendix) gives an account of the distribution of ranks of the initial placements in the final sample.[16]

Our analysis focuses on academic placements only. Out of the 3,182 students in our sample, 1,372 (roughly 43%) initially went to a ranked university or college during the four academic years under consideration. Students interested in a research career may not only go to universities, but to the private, governmental and non-governmental sector. For example, 31 students were hired by research institutes, 7 to policy institutes and 5 to military organisations (US Air Force Academy, US Naval Academy, etc.). Other big players in the market are research-active financial institutions such as the International Monetary Fund (75 PhD students), the World Bank (44 PhD students), the Bank for International Settlements (4 PhD students) and various central banks (87 PhD students). We also note that the academic career is not the only possible career fresh Economics graduates want to pursue (Stephan, 2012; Conti and Visentin, 2015). Research intensive private companies (Cornerstone Research etc.) hired 25 students, while consulting firms hired another 22 students.

We augment our data with information about the 'tightness' of yearly job markets. We use yearly "Reports of the Director Job Openings for Economists" published in The American Economic Review,[17] which give an account of the number of openings per field (as defined by

[14] For details see `https://econtop.uvt.nl/methodology.php`.

[15] For the original ranking use for various years `https://econtop.uvt.nl/`. Spearman correlation between our version and the original for 2004 is 0.89.

[16] Estimations using the unweighted count of publications in these 70 journals do not alter the qualitative results. The likely reason is that the 70 Economics journals used to measure the weighted publication output are all very good journals.

[17] In particular, we use the reports by Hinshaw (2000, 2001, 2002, 2003) and Hinshaw (2004).

the JEL code). To measure tightness in a given year for a given field, we divide the supply of students in a field by the number of openings in that field:

$$\text{Tightness}_{tf} = \frac{Students_{zf}}{\text{Openings}_{tf}} \tag{4.1}$$

where z is the academic year starting in the second half of year t and f is the JEL code-defined field. That is, students finishing in the first half of a year are accounted to the previous year's job market. Clearly, a higher tightness ($Tightness_{tf}$) would indicate that the market is more 'difficult' for students in field f in year t (higher supply of students in that field compared to demand). In our analysis, we show that adviser connectedness matters more for her student's placement in years with higher tightness as compared to years with lower tightness.

4.3.3 Networks of Collaboration

Our variable of interest is adviser connectedness in the Economics co-author network. In a co-author network, nodes represent researchers, and a link exists between two nodes if the researchers have jointly published a full research article. Co-author networks have sparked great interest among Economists, starting with Eagly (1975) who describes "Economics Journals as Communications Network". More recently Goyal et al. (2006) have shown that Economics co-author networks since the 1990s have small-world properties, implying that communication is greatly facilitated by a few highly interlinked stars. Ductor et al. (2014) show that one's current local network has predictive value for one's future productivity.

To construct the co-author networks we consider 114,409 publications indexed in Elsevier's Scopus database published in 408 journals between 1997 and 2005. Since our analysis covers the period up until 2004, we include co-author ties visible one year later (as the research project must have begun earlier). The set of journals from which we draw our co-author network is defined according to field-wise rankings in Combes and Linnemer (2010). We include every title that is ranked at least C in any field-wise ranking. In this network two researchers are connected when they have jointly published a paper, where the link weight corresponds to the number of joint publications. The idea behind the weight is that if a pair of researchers has published several papers as co-authors then they have a stronger connection compared to if they had worked together only once.

For each year $t (\in \{2000, 2001, 2002, 2003, 2004\})$ we construct the network using the publications published in all years $1996, 1997, \cdots, t, t+1$. We chose the network definition such that network variations comes from new connections and old connections are not disregarded. As the number of articles increased over time, the network grew too. In the earliest of our networks, the one for the year 2000, there are 30,617 distinct researchers. The network for 2004 consists of 52,942 distinct researchers. These networks are represented by symmetric matrices G whose entries g_{mn} indicate the strength of a link between m and n. The diagonal is set to 0.

For technical reasons we only consider the network's giant component. This is the network's largest component where each node is accessible from any other node by an uninterrupted series of links. Two nodes are said to be in two different components when there is no such path of links. While it is theoretically possible to compute centralities for each component, they are not comparable, as the computation takes into account the size of each component. The respective giant component for our analysis covers about one third of the overall network size.

Eigenvector centrality is a measure of influence and defined as the weighted sum of the Eigenvector centralities of the network neighbors, where the weight corresponds to the neighbor's own Eigenvector centrality. The idea is that if one is connected to nodes that are themselves more connected then one is more connected. The centrality score is obtained as a fixed point that satisfies, for scalar λ and any non-zero vector $\mathbf{E}$:

$$\mathbf{E}G = \lambda \mathbf{E}. \tag{4.2}$$

By the Perron-Frobenius theorem, if equation (4.2) holds then λ is the leading eigenvalue of G, called $\mu_1(G)$, and vector $\mathbf{E}$ the assocciated eigenvector. The elements of $\mathbf{E}$ are hence the eigenvector centralities of all members in G.

A crucial element in the calculation of author centralities is that we remove the adviser from the network before we compute the adviser's coauthors' Eigenvector centrality. The reason is that our identification strategy exploits *exogenous* changes to the adviser's centrality coming from changes in the centrality of her co-authors. Clearly, we would not want these changes to result from changes in the centrality of the adviser herself. Formally, we refer to the network as square matrix G, and to the network without adviser a of student i as G_{-a_i}:

$$\mathrm{E}(m) = \frac{1}{\mu_1(G_{-a_i})} \sum_{n \in N_m} g_{mn} \mathrm{E}(n), \tag{4.3}$$

where N_m is the set of co-authors of m.

Finally, since the centrality scores only indicate the relative importance of different nodes, we convert the scores into ranks for nodes. This also helps because ranks make centrality positions more comparable across networks for different years. Our variable of interest is hence the average Eigenvector centrality rank of the coauthors N_{a_i} of adviser a of student i in t:

$$\text{Adv. neigh. mean Eigenvector rank}(i,t) = \sum_{o \in N_{t+1,a_i}} \frac{\text{rank}\big(E(o, G_{t+1,-a_i})\big)}{|N_{a_i}|} \tag{4.4}$$

In the second part of our analysis we are interested in the connectedness of advisers with universities, rather than with other economists. We use the Hasselback Faculty Directories for Economics, Management and Finance to obtain information on faculty membership[18]. Faculty rosters for Economics exists for 2001/2002 and 2003/2004 academic years, for Management for the 2001/2002 academic year, and for Finance for the academic years 2000/2001, 2002/2003 and 2004/2005. The rosters include 14105 distinct faculty members which we could identify on Scopus (a pre-requirement to be in the co-author network). 6040 of the faculty members are also nodes in the co-author network.

[18] See http://www.jrhasselback.com/FacDir.html. The lists are sometimes called Prentice Hall Guide to Economics Faculty resp. Prentice Hall Guide to Finance Faculty individually.

4.4 Empirical Issues and Identification Strategy

Our objective is to identify the effect of adviser connectedness on student placement. Additionally, we would like to provide supportive evidence to show that the channel through which adviser connectedness helps students on the job market is by reducing information asymmetry about the student's quality (via informal phone calls/emails). There are several endogeneity issues we need to contend with.

Consider the problem of identifying the impact of the adviser's connectedness on student placement (we discuss issues with identifying the channel through which connectedness affects placement in section 4.6). An adviser's connectedness in the network of co-authors is not exogenous: advisers who are more productive, are more experience or affiliated with better universities are more likely to be better connected. Therefore, a simple regression of student placement on adviser connectedness may pick up the impact of these variables rather than that of connectedness. Furthermore, while we can mitigate the above effects by controlling for the adviser's publication record, university and seniority (which we do), there might be unobserved variables which affect both adviser connectedness and student placement. Ideally, we would like a variable which exogenously shifts adviser connectedness but does not affect student placement through any other channel.

Our strategy is to exploit the longitudinal nature of our placement data and identify the impact of adviser connectedness on student placement via *changes* in the connectedness of the adviser's co authors in a model with adviser fixed effects. The connectedness of an adviser's co-authors' can change when her co-authors start new projects with new co-authors. This affects the centrality score of the adviser as well. To avoid having changes in the adviser's own centrality causing the change in her co-authors' centrality, we compute the centralities of the adviser's co-authors' in the network of co-authors which excludes the adviser herself. While the centrality *level* of an adviser's co-authors is endogenously determined, our identifying assumption is that the *change* in the connectedness of an adviser's co-authors *in the year of placement of the student* is not anticipated, and is therefore not strategically used by either the adviser or her students for better placement. Thus, this variable does not directly affect the placement of the adviser's students except to the extent that it changes the adviser's connectedness in the year in which the student is on the job market. We include adviser-fixed effects in our regressions to make sure that there are no unobserved time-invariant adviser characteristics biasing our coefficients. This strategy works as an IV for all advisers who place graduate students in more than one year in our data set. Hence, for this sample we exclude students whose advisers did not place another student in another year. We run the regressions using the mean centrality rank of the adviser's co-authors as the main explanatory variable.

Additionally, we show results without adviser-fixed effects, since this allows us to use more data. The identification assumption is stronger in this case though. The underlying identifying assumption in the model without adviser fixed effects is that after controlling for the adviser's publication record, experience, gender and affiliation, the only channel through which the *level* of connectedness[19] of an adviser's co-authors' affects student placement is by affecting the connectedness of the adviser.

Now, we discuss more deeply the specific channels which could bias our results, and how our identification strategy assuages these concerns. First, there are many unobserved characteristics of the adviser which could be correlated with both adviser connectedness (even with changes in adviser connectedness) and student placement. For example, one may argue that better/smarter advisors are more likely to increase their network in any given year via new collaborations. Or, a different channel could be that more 'helpful' advisers are more likely to write papers with their students, and younger economists are likely to engage in more collaborative projects. Thus, being helpful may affect the change in connectedness of any adviser (via changes in the centrality of their ex student-coauthors), and also affect their student's placement in any given year. We address all such concerns about unobserved adviser characteristics biasing our coefficients by including adviser fixed effects in our models. Additionally, we include controls for several observed adviser characteristics like publication quality and experience which are not time invariant, and could influence student placement.

Next, we discuss if unobserved student quality and assortative matching could bias our

[19]Note that this considers only the level of connectedness (in the model without adviser fixed effects) of the adviser's co-authors, not the change in the level of connectedness since we don't have adviser fixed effects.

results. We don't have sufficient controls for an important variable which affects student placement - student quality. This could bias our estimates if good students are more likely to match with more connected advisers. If this were the case then good placements will be because of high student quality and not because of the connectedness of the adviser. For this issue, we have the following argument. Our identification strategy breaks down only if there is an unobserved variable which affects student-adviser matches and is simultaneously correlated with the change in the centrality of the adviser's co-authors in the year in which the student gets placed. However, we believe that it is hard for students (and advisers) to anticipate the change in connectedness[20] of their adviser's co-authors' *in the year of their placement*. Thus, we argue that the change in the centrality of the adviser's co-authors affects student placement only via (exogenously) changing the adviser's connectedness.

Another counter argument could be that of clustering. The concern with clustering is that *helpful advisers have helpful co-authors and helpful co-authors are more likely to have changes in connectedness (because more people are willing to work with them)*. However, the impact of clustering must die down with distance from the adviser. That is, the probability of sharing a quality reduces with distance. As a robustness check we look at the co-authors of an adviser's co-authors (i.e. the second neighbors of the adviser) and then take their connectedness as our IV for adviser connectedness in a model with adviser-fixed effects.

Before we move on to describe a data issue we face, we would like to point out that despite our best efforts there are some remaining channels which may bias our results. We leave it to the judgment of the reader to determine how big these effects may be. One example of such a channel would be if a subfield suddenly became popular. If both the adviser and her student work in this subfield, then this will affect both the change in the connectedness of the adviser (more papers will be written in this subfield which will mean more collaborations), and the ranking of the student's placement (more universities may be interested in hiring in the newly popular subfield). We wish to make two points regarding this channel. One, we control for field fixed effects so the above channel can only work for smaller subfields. Two, for the above

[20] Students may know the approximate *level* of their adviser's connectedness in the year in which they choose their adviser but it would be quite difficult to anticipate the precise *change* in connectedness of their adviser's co-authors in their expected future graduation year.

argument to bias our coefficients, the subfield would have to suddenly become popular *within* the short time period of our data set (1999-2004).

A significant data issue we face is that of sample selection. Due to data limitations and our identifying restrictions, we lose data points from the initial sample size of 3,182 graduates. The main restrictions are - a) we can only consider students who were placed in Tilburg ranked departments, b) we can only consider those students whose adviser placed students at Tilburg ranked departments in multiple years between 2000 and 2004 (to employ adviser fixed effects), and c) we study only those students for whom we have both placement and adviser information. The final sample size is 366 students (about 11.5 percent of the original sample). Losing a large fraction of our student data set can lead to selection bias in our estimated coefficients. While we have no strong arguments to negate the effect of this data issue, we would like to make the following observation about the probable direction of bias. Table 4.2 shows that the final sample size is biased towards higher-ranked institutions: about 47% of all students in the initial dataset (of 3,182 students) received their PhD from a university ranked 30 or better in the year of their graduation. This share increases to about 70% in the final sample size we use for our regressions. We expect that adviser connectedness matters more for students who are not from top schools. This is because students from better ranked universities usually reduce information asymmetry about their quality via publications before the job market, and by presenting their work at top conferences (this also gives them a chance to network). Therefore, it is possible that our estimate of the impact of adviser connectedness on student placement is a conservative one.

Finally, we also show that co-author networks matter for student placement in a simpler way. Suppose the distance between a department and an adviser is defined as the shortest path in the co-author network between the adviser and any faculty member at the department. We use the death of economists (anywhere in the co-author network) as an exogenous shock which affects the social distance between an adviser and different departments negatively.[21] We then ask whether an increase in the social distance due the death of economists affects the

[21] The death of an economist could increase the distance between an adviser and a department by breaking the shortest co-author path the adviser had to that department.

Table 4.2: Comparison of initial and final dataset.

	Initial		Final	
	Number	Share (in %)	Number	Share (in %)
Ranks 1-30	1458	45.82	257	70.22
Ranks 31-100	799	25.11	88	24.04
Ranks 101-300	718	22.56	17	4.64
Other	207	6.51	4	1.09

Notes: Table lists number of students and share of total by PhD school group for the initial and the final dataset. PhD schools are grouped according to the rank in the student's year of graduation. Students of 6 PhD schools without rank in the year of the graduation were categorized together with "Other".

probability of the adviser's student getting placed in the department. Examining the effect of deceased individuals on their local network is a popular identification strategy in the study of social networks.(Azoulay et al., 2010; Fracassi and Tate, 2012; Oettl, 2012; Azoulay et al., 2015) There are 30 researchers who passed away during the 1999-2004 period.[22] Table 4.A.3 in the appendix gives an account of these researchers along with their date of death.

4.5 Empirical results

4.5.1 Centrality rank and placement rank

We look at all those students in our sample for which the following three requirements hold: (1) the student was placed at a ranked institution, (2) the student's adviser is a member of the co-author network's giant component, (3) conditions (1) and (2) hold for at least one other student of the same adviser who graduated in a different placement year. Condition (3) is crucial for our identification strategy since want to use a model with adviser-fixed effects to obtain identification via *changes* in the centrality of the adviser's co-authors. Our final sample is called the 'adviser coauthor centrality sample', and it consists of 366 students.

[22]We let the period start one year earlier than our student sample because vacancies from researchers that passed away in the year are likely to not be filled up so soon.

For this sample, the summary statistics are presented in table 4.3, while table 4.A.2 (in the appendix) reports the relevant correlation coefficients. The mean placement in the year of the placement has rank 256.7, which for 2002 refers to the Economics departments of the Universities of Valencia, Complutense de Madrid, Adelaide, Trinity College Dublin, Syddansk Universitet, Binghamton, West Virginia, Eindhoven (TU), Vilnius and Lausanne, as well as American University, Indiana University-Purdue University Indianapolis, and Queen's University Belfast. It ranges from rank 1 (Harvard University, in 2004) to rank $1,416$. Figure 4.A.2 in the appendix presents the distribution graphically. Our variable of interest is the adviser's Eigenvector centrality rank and this has a mean of 6149.9, ranging from 240 to $27,629$.[23] The instrument for this variable is the adiser's coauthors' mean Eigenvector centrality rank, which is computed in the network without the adviser. Its means is 7638.9. The average Euclidean Index of citations for the adviser is 245.6[24] and ranges from 1 to 2530. Adviser experience is measured by the number of years since the first indexed publication. This is 18.8 years on average, ranging from 3 to 45 years.

For our analysis, we estimate the following regression equation in an ordered logistic regression model[25]:

$$\begin{aligned}\text{PlacementRank}_{it} =& \beta_0 + \beta_1 \text{AdvisersCoauthorsMeanEigenvectorRank}_{it} + \\ & \beta_2 \text{Gender}_i + \beta_3 \text{PhDSchoolRank}_{it} + \beta_4 \text{AdviserControls}_{it} + \\ & \gamma_1 \text{Adviser}_i + \gamma_3 \text{YearOfCompletion}_i + \gamma_4 \text{Field}_i + \epsilon_{it} \end{aligned} \tag{4.5}$$

The outcome variable is the placement rank of student i in year t, where the rank is computed following the methodology of the Tilburg University Economics ranking but using data from Scopus. $AdvisersCoauthorsMeanEigenvectorRank_{it}$ is the mean Eigenvector centrality rank over i's adviser's coauthors in the weighted coauthor network for year t, computed in a network without the adviser (equation (4.4)). We are interested in β_1. Note that since lower

[23] The minimum rank equals 240 and not 1 as one might expect for two reasons. First, it's the average over all neighbor's centrality ranks, and secondly our ranks are relative to the entire population that makes up the network.

[24] A possible paper/citation trajectory would be 74, 43, 10 and 230.

[25] Results of an ordered probit regression are qualitatively the same.

Table 4.3: Summary statistics in the adviser coauthor centrality sample.

	N	Mean	Median	Std.Dev.	Min	Max
Placement Rank	363	256.7	97	355.00	1	1416
Adv. Eigenvector rank	567	6149.9	4964	4384.66	240	27629
Adv. neigh. mean Eigenvector rank	567	7638.9	7261	3761.34	269	24390
Tightness	573	0.304	0.282	0.167	0.08	1.048
School Rank	574	33.5	14	56.56	1	521
Adv. Euclidean Index	574	245.6	116	386.26	2	2530
Adv. Experience	574	18.8	18	7.62	3	45

Notes: Placement Rank is the rank according to our version of the Tilburg University Economics ranking of a student's placement in the year of the placement. *Adv. neigh. mean Eigenvector rank* is the mean Eigenvector centrality rank of all coauthors of an adviser in the weighted coauthor network corresponding to the year of the placement. *Market tightness* for a field in a given year is the number of students graduating in that year in that field, divided by the number of AEA-reported job openings in that field in that year, where field is measured by JEL code (equation (4.1)). *PhD School Rank* is the rank according to our version of the Tilburg University Economics ranking of the PhD-awarding university in the year the student finished. *Euclidean Index* is the adviser's Euclidean index of citations in the year the student graduated. *Experience* is the number of years between an adviser's first publication and the year in which the student graduated.

numbers indicate better ranks[26], we expect β_1 to be positive as this would indicate a positive relationship between adviser centrality and better placement. $Gender_i$ is a binary variable indicating a female student. We also have another gender control (*samesex*) which is a binary variable indicating that student and adviser have the same sex. $AdviserControls_{it}$ include the adviser's Euclidean index of citations in t, her experience and experience squared, because these values are time-variant. $PhDSchoolRank_{it}$ is the rank according to our version of the Tilburg University Economics ranking of the PhD granting school of student i in year t. In a variation of the model we replace $PhDSchoolRank_{it}$ with university-fixed effects. In all specifications we cluster standard errors at the PhD granting University level to allow for unobserved heterogeneity as well as different group sizes.[27] Fixed-effects for the year of completion captures year-specific information. We also include fixed-effects for the student's field. Adviser-fixed effects

[26] Both the dependent variable $PlacementRank_{it}$ and the main explanatory variable $AdviserCoauthorMeanEigenvectorRank_{it}$ are ranks.

[27] Abadie et al. (2017) argue that standard errors must be clustered around a variable when there is selection bias in the sample on that variable. We definitely get more students from better ranked universities in our final sample (see Table 4.2).

control for unobserved adviser characteristics which could influence student i's placement.

Table 4.4 presents results from an ordered logit regression model (regression equation (4.5)). In column (1), the adviser's neighbor's mean Eigenvector centrality rank is statistically insignificant with a p-value of 0.121 but has the expected sign. In column (2) we use the take the mean of the three most-connected co-authors of an adviser. The coefficient is statistically significant with a p-value equal to 0.029. Column (3) follows the specification of column (1), except that we include PhD School-fixed effects instead of adviser-fixed effects. The coefficient is statistically significant with a p-value equal to 0.055.

To given an idea of the effect size, we take the coefficient of model (2). An improvement in the adviser's three most connected co-authors' average Eigenvector centrality rank by 1 results in a $e^{0.0000697} - 1 \approx 0.007\%$ better placement outcome, where 'better' refers to a better ranked department according to our version of the Tilburg University Economics ranking. A 0.007% improvement translates into $0.00007 * 1416 \approx 0.1$ placement ranks. To put this into perspective, there are between 26,000 and 40,000 nodes in the yearly varying co-author networks, as nearly every researcher has a unique rank. Hence, the scope for rank change among an adviser's coauthors' is considerably large. An improvement in 10 ranks in the mean rank of an adviser's three most connected coauthors would result in a placement rank improvement of the adviser's student by approximately 1.

The coefficient for adviser-student relationships of same sex is positive but statistically insignificant, whereas the coefficient for female students is positive and weakly statistically significant. Female students tend to be placed worse, and the effect is partly offset in magnitude if their adviser is female, too. However, within a PhD school this effect disappears (column (3)). This is interesting in light of recent studies investigating same sex-relationships in PhD student advising (Gaulé and Piacentini, 2017; Mansour et al., 2018).

We hypothesize that if informal contacts do matter for placements, then their value should increase with tighter market conditions. That is, the adviser's connectedness should play a stronger role in student placement in the years in which there is more competition for placement. To test this hypothesis we interact our variable of interest, the adviser's coauthors' Eigenvector centrality rank with market tightness. Table 4.5 presents regression results in a model where are our variable of interest is interacted with market tightness as defined in equation (4.1). As before, we include adviser-fixed effects, field-fixed effects and cluster standard errors around the PhD school. Unlike before, we do not control for placement year. This is because if we include year fixed effects then field and year-fixed effects would completely absorb any market tightness variation. In column (2) we replace school rank with school-fixed effects, and in column (3) we drop adviser-fixed effects to gain more observations. In the first two models, the interaction of Adv. neigh. mean Eigenvector rank and Market Tightness is statistically significant and has a negative sign. Thus, we confirm our hypothesis that an adviser's connectedness matters more when market conditions are tight.

A possible concern with our centrality regression is that of 'clustering'. We use changes in the centrality of the adviser's co-authors to identify the impact of the adviser's centrality on the placement of her students. However, this relies on the assumption that changes in the centrality of the adviser's coauthors only influence the placement of the adviser's students by changing the centrality of the adviser. This will not be true if the aforementioned variable is correlated with an unobserved adviser characteristic which also affects placement. Consider for example a helpful adviser's who could have helpful co-authors who are more likely to have changes to their connectedness because many people want to work with them. As a robustness check against this kind of argument, in table 4.6, we report regression results repeating the above analysis but using the average centrality rank of the adviser's *second* (or indirect) neighbors. The idea is that the clustering of unobserved characteristics must die with distance. An adviser's second neighbors are less likely to share the adviser's unobserved characteristic. We show that the placement of the adviser's students is still significantly affected by the centrality of the adviser's second neighbors. This also strengthens our argument for the importance of even indirect connections in job placement. Columns (1), (2), (3) and (4) include adviser-fixed effects, while columns (5) and (6) replace adviser-fixed effects PhD school-fixed effects. Columns (3), (4) and (6) interact our variable of interest with market tightness and thus exclude year-fixed effects.

Table 4.4: Results of Ordered logistic regression for rank of initial placement, adviser coauthor centrality sample.

	Placement Rank		
	(1)	(2)	(3)
Adv. neigh. mean Eigenvector rank	0.0000584 $p = 0.121$		0.0000461* $p = 0.055$
Adv. top 3 neigh. mean Eigenvector rank		0.0000697** $p = 0.029$	
Same sex	0.579 $p = 0.209$	0.613 $p = 0.233$	−0.227 $p = 0.524$
Female student	0.845* $p = 0.057$	0.876* $p = 0.076$	−0.0820 $p = 0.825$
PhD School Rank	0.00982*** $p = 0.002$	0.00910*** $p = 0.006$	
Euclidean Index	−0.000753 $p = 0.242$	−0.000742 $p = 0.208$	−0.0000772 $p = 0.881$
Experience	−0.0534 $p = 0.468$	−0.0606 $p = 0.402$	−0.0332 $p = 0.404$
Experience^2	0.00148 $p = 0.329$	0.00168 $p = 0.264$	0.00137 $p = 0.157$
Euclidean Index (2014)			−0.000139 $p = 0.158$
Adviser fixed effects	*Yes*	*Yes*	*No*
PhD School fixed effects	*No*	*No*	*Yes*
Field fixed effects	*Yes*	*Yes*	*Yes*
Year Fixed effects	*Yes*	*Yes*	*Yes*
Clustered SE	PhD School	PhD School	PhD School
N	357	353	692

Notes: ***, ** and * indicate statistical significance to the 1, 5 and 10 percent level, respectively. *Placement Rank* is the rank according to our version of the Tilburg University Economics ranking of the student's placement in the year of the placement. *Adv. neigh. mean Eigenvector rank* is the mean Eigenvector centrality rank of all coauthors of a student's adviser in the weighted coauthor network in the year of the student's placement, computed without the adviser (equation (4.4)). Similarly, *Adv. top 3 neigh. mean Eigenvector rank* is the mean Eigenvector centrality rank of an adviser's three most connected coauthors. *Same sex* equals 1 if both the student and the adviser are estimated to be of the same sex. *Female student* equals 1 if the student's first name is estimated to be a female first name. *PhD School Rank* is the rank according to our version of the Tilburg University Economics ranking of the PhD-awarding university in the year the student finished. *Euclidean Index* is the adviser's Euclidean index of citations in the year the student graduated. *Experience* is the number of years between an adviser's first publication and the year in which the student graduated. *Experience*^{2} is its square. *Euclidean Index (2014)* is the adviser's Euclidean index not in the year of student placement, but in 2014.

Table 4.5: Results of an Ordered Logistic regression for rank of initial placement interacted with market tightness, adviser coauthor centrality sample.

	Placement Rank		
	(1)	(2)	(3)
Adv. neigh. mean Eigenvector rank	−0.0000408 0.573		−0.0000345 0.532
Market Tightness	-4.783^{**} 0.049	-3.091^{**} 0.035	−2.956 0.218
Adv. neigh. mean Eigenvector rank × Market Tightness	0.000371^{*} 0.100		0.000327^{*} 0.074
Same sex	0.433 0.390	0.404 0.486	−0.228 0.520
Female student	0.736 0.106	0.742 0.166	−0.0858 0.814
PhD School Rank	0.00971^{***} 0.004	0.00913^{**} 0.013	
Euclidean Index	−0.000691 0.280	−0.000624 0.275	0.000194 0.713
Experience	−0.0339 0.627	−0.0275 0.679	−0.0340 0.398
Experience square	0.000960 0.490	0.000864 0.519	0.00125 0.189
Adv. top 3 neigh. mean Eigenvector rank		−0.0000433 0.332	
Adv. top 3 neigh. mean Eigenvector rank× Market Tightness		0.000379^{***} 0.007	
Euclidean Index (2014)			-0.000197^{*} 0.062
Adviser-fixed effects	*Yes*	*Yes*	*No*
PhD School-fixed effects	*No*	*No*	*Yes*
Field-fixed effects	*Yes*	*Yes*	*Yes*
Clustered SE	PhD School	PhD School	PhD School
N	357	353	684

Notes: ***, ** and * indicate statistical significance to the 1, 5 and 10 percent level, respectively. *Placement Rank* is the rank according to our version of the Tilburg University Economics ranking of the student's placement in the year of the placement. *Adv. neigh. mean Eigenvector rank* is the mean Eigenvector centrality rank of all coauthors of the student's adviser in the weighted coauthor network in the year of the student's placement, computed in a network without the adviser (equation (4.4)). *Female student* equals 1 if the student's first name is estimated to be a female first name. *Same sex* equals 1 if both the student and the adviser are estimated to be of the same sex. *Market tightness* for a field in a given year is the number of students graduating in that year in that field, divided by the number of AEA-reported job openings in that field in that year, where field is measured by JEL code (equation (4.1)). *PhD School Rank* is the rank according to our version of the Tilburg University Economics ranking of the student's PhD granting institution in the year of the student's graduation. *Euclidean Index* is the adviser's Euclidean index of citations in the year the student graduated. *Experience* is the number of years between an adviser's first publication and the year in which the student graduated. *Experience*^{2} is its square. *Adv. top 3 neigh. mean Eigenvector rank* is the mean Eigenvector centrality rank of an adviser's three most connected coauthors. *Euclidean Index (2014)* is the adviser's Euclidean index not in the year of student placement, but in 2014.

Table 4.6: Results of an Ordered logistic regression for rank of initial placement for second neighbor centrality ranks, adviser coauthor centrality sample.

	Placement Rank					
	(1)	(2)	(3)	(4)	(5)	(6)
Adv. 2nd neigh. mean Eigenvector rank	0.0000663* $p = 0.070$		−0.0000192 $p = 0.790$		0.0000571* $p = 0.059$	−0.0000234 $p = 0.661$
Adv. top 3 2nd neigh. mean Eigenvector rank		0.0000745** $p = 0.043$		−0.0000767 $p = 0.113$		
Same sex	0.567 $p = 0.230$	0.580 $p = 0.238$	0.349 $p = 0.515$	0.286 $p = 0.639$	−0.215 $p = 0.544$	−0.222 $p = 0.539$
Female student	0.844* $p = 0.059$	0.831* $p = 0.071$	0.686 $p = 0.151$	0.620 $p = 0.266$	−0.0747 $p = 0.839$	−0.0797 $p = 0.829$
PhD School Rank	0.00977*** $p = 0.002$	0.00960*** $p = 0.003$	0.00973*** $p = 0.003$	0.00997*** $p = 0.007$		
Euclidean Index	−0.000810 $p = 0.202$	−0.000724 $p = 0.272$	−0.000749 $p = 0.236$	−0.000647 $p = 0.268$	−0.000151 $p = 0.770$	0.000103 $p = 0.842$
Experience	−0.0477 $p = 0.505$	−0.0528 $p = 0.443$	−0.0239 $p = 0.722$	−0.0184 $p = 0.767$	−0.0297 $p = 0.467$	−0.0290 $p = 0.474$
Experience square	0.00142 $p = 0.337$	0.00151 $p = 0.286$	0.000845 $p = 0.531$	0.000648 $p = 0.605$	0.00131 $p = 0.192$	0.00116 $p = 0.238$
Market Tightness			−4.361* $p = 0.071$	−2.142 $p = 0.123$		−3.034 $p = 0.210$
Adv. 2nd neigh. mean Eigenvector rank × Market Tightness			0.000335 $p = 0.134$			0.000332* $p = 0.064$
Adv. top 3 2nd neigh. mean Eigenvector rank × Market Tightness				0.000482*** $p = 0.001$		
Adviser fixed effect	*Yes*	*Yes*	*Yes*	*Yes*	*No*	*No*
PhD School fixed effects	*No*	*No*	*No*	*No*	*Yes*	*Yes*
Field fixed effects	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Year fixed effects	*Yes*	*Yes*	*No*	*No*	*Yes*	*No*
N	357	357	357	357	692	691

Notes: ***, ** and * indicate statistical significance to the 1, 5 and 10 percent level, respectively. *Placement Rank* is the rank according to our version of the Tilburg University Economics ranking of the student's placement in the year of the placement. *Adv. 2nd neigh. mean Eigenvector rank* is the mean Eigenvector centrality rank of all indirect coauthors of the student's adviser in the weighted coauthor network in the year of the student's placement, computed in a network without the adviser. *Same sex* equals 1 if both the student and the adviser are estimated to be of the same sex. *Female student* equals 1 if the student's first name is estimated to be a female first name. *PhD School Rank* is the rank according to our version of the Tilburg University Economics ranking of the student's PhD granting institution in the year of the student's graduation. *Market tightness* for a field in a given year is the number of students graduating in that year in that field, divided by the number of AEA-reported job openings in that field in that year, where field is measured by JEL code (equation (4.1)). *Euclidean Index* is the adviser's Euclidean index of citations in the year the student graduated. *Experience* is the number of years between an adviser's first publication and the year in which the student graduated. *Experience^2* is its square.

4.5.2 Social Distance and Placement probability

In this subsection, we present additional evidence that an adviser's connections in the academic network matter for her student's placement. We show that the social distance between an adviser and a department affects the probability of the adviser placing her student at that department.

Hitherto, we have focused on identifying the impact of adviser centrality on the rank of her student's placement. In the data sample for this subsection, which we term "adviser distance sample", the unit of observation is the connection between advisers and placement university's faculty members. We look at all possible paths between every adviser and every possible university that satisfies three conditions: a) there is a department rank available for the university or college, b) we know the faculty members from the Hasselback rosters and c) at least one faculty member is in the co-author network. We then count the number of steps one has to take to go from an adviser a to the closest member of university k (where each step is a co-author link). Thus, we define the distance between an adviser a and a university k by the length of the shortest path. In order to identify the impact of this 'social distance' between an adviser and a prospective placement university on the placement of the adviser's student, we construct a variable that measures the increase in social distance caused by the death of authors somewhere in the network. The dependent variable is whether one of a's students were placed at university k in t.

We estimate the following regression equation in a logistic regression model[28]:

$$\begin{aligned} Placement_{akt} =& \beta_0 + \beta_1 \text{IncreaseInSocialDistanceAfterDeath}_{akt} + \beta_2 \text{SocialDistanceBeforeDeath}_{akt} \\ &+ \gamma_1 \text{PlacementRank}_{kt} + a + \text{PhDSchool}_j + t + \epsilon_{jkt} \end{aligned} \tag{4.6}$$

We are interested in coefficients β_1 and β_2. Since the social distance increases due to the removal of deceased authors, we expect β_1 to be negative, as this indicates a lower probability of student placement at k. The variable *SocialDistanceBeforeDeath*$_{akt}$ indicates the length of the

[28]Results of a probit model are qualitatively the same.

shortest path between adviser a to the nearest faculty member of k in the co-author network in year t, given that the path exists (before accounting for the change in distance due to death of authors). We expect this coefficient to be negative (though not identified) since a shorter distance to another faculty should result in a higher placement probability if social connections do play a role in placement. $PlacementRank_{kt}$ is the rank according to our version of the Tilburg University Economics ranking of university k in year t. Fixed effects for adviser a and the student's Phd granting school $PhDSchool_j$ capture unobserved characteristics, while fixed effects for year of placement capture market characteristics for that year. In a variant of the model we remove adviser-fixed effects to gain more observations, and instead we control for adviser characteristics such as her Euclidean Index of citations, her experience and the squared experience. In all variants of the model we cluster standard errors around *PhDSchool* to account for unobserved heterogeneity and different group sizes.

Table 4.7: Summary statistics for adviser distance sample.

	N	Mean	Median	Std.Dev.	Min	Max
Student placement	1027	1.00	1	0.00	1	1
Increase in social dist. after death	46910	1.50	1	0.86	1	11
Social dist. before death	973294	8.72	8	2.74	2	32
Placement Rank	973294	187.82	98	293.89	0	2554
Euclidean Index	973294	18.84	18	8.41	0	52
Experience	973294	263.87	215	214.55	1	845
Male adv.	973294	0.91	1	0.29	0	1

Notes: All variables refer to time-variant dyads between adviser a and placement k in year t, given that k appears in our ranking and a list of faculty members is available. *Student Placement* equals 1 if a student of adviser a was placed at university k in year t. *Increase in social dist. after death* is the increase in social distance in the co-author network between a and the nearest faculty member of k after scientist died in year $t-1$. *Social dist. before death* is the social distance in the co-author network between a and the nearest faculty member of k before the distance changed induced by the removal of deceased scientists. *Placement Rank* is the rank according to our version of the Tilburg University Economics ranking of university k in year t. *Euclidean Index* is the adviser's Euclidean index of citations in year t. *Experience* is the number of years between an adviser's first publication and year t. *Male adv.* equals 1 if the adviser is estimated to not be female. Only paths between known advisers of students graduating from North-American universities and identified faculty members of departments listed in our version of the Tilburg University Economics ranking and listing in the Hasselback faculty roosters considered.

Table 4.7 presents summary statistics for the adviser distance sample. The total number of dyads (adviser-university pairs) is $2,706,900$, of which $973,294$ include existing paths between adviser a and the closest faculty member of k in year t. For $46,910$ of these, the social distance increased due to the exogenous shock of author deaths, sometimes up to 11 steps. Corresponding correlation coefficients are given in table 4.A.4 (in the appendix).

Table 4.8 presents results of a logistic regression for model 4.6 with standard errors clustered around PhD School. As expected, the coefficient of social distance in the co-author network before death is negative, indicating that students are more likely to be placed at faculties to which their adviser has a shorter distance. A one unit increase in the distance to university k decreases the odds of being placed at k by $e^{-0.363} - 1 \approx -30\%$ (at the mean co-author distance before death (1.5) and holding all other variables fixed at the mean). The coefficients change marginally in column (2), where we control for the PhD School rather than the adviser. The p values for both models fluctuate around a value of 0.06, indicating robust statistical significance.

4.6 Channel

We want to argue that an adviser's connectedness in the co-author network matters for her student's placement because it reduces information asymmetry regarding the match quality of a student with a university. Other theoretically possible channels include reciprocity and favoritism. We will provide supporting empirical analysis (without showing any causality) to argue that the latter channels are not very important.

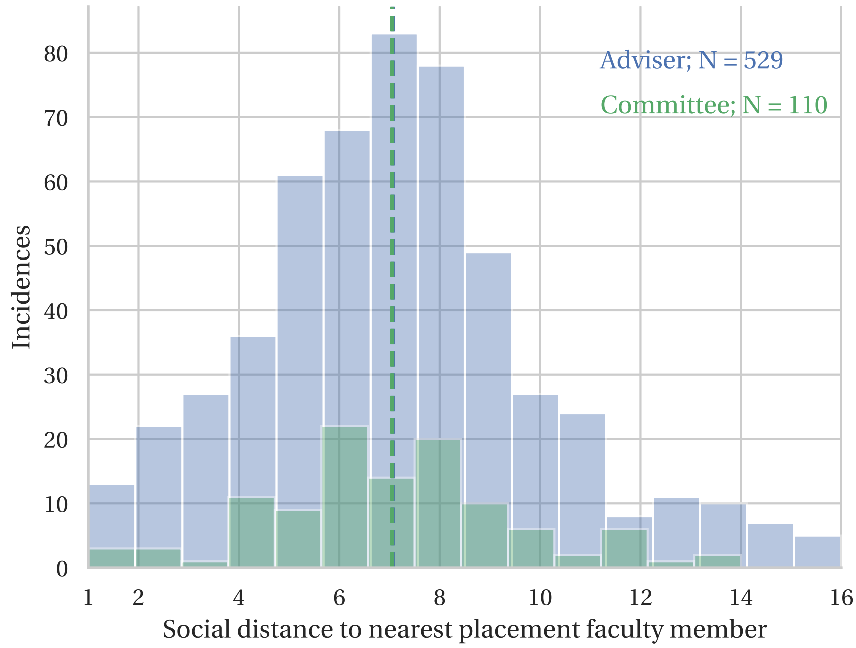
Reciprocity refers to a direct exchange of students in the same or subsequent years between two universities ('*I hire your students, you hire mine*'). More connected advisers may have more of these reciprocal relationships which would affect student placement. To assess the importance of this channel we construct a network of PhD granting schools connected via directed links whenever one school places a PhD student in the other. There are 437 non-zero links between universities in our sample (i.e. those that produce PhD students) of which only

5.1% of all ties are bi-directional i.e. both universities hired a student from the other (indicating an exchange of students). Being very rare, bidirectional links tend to occur more often with top schools, i.a. a student moving from Harvard to MIT and vice verse. Based on the low number, we rule out reciprocity as an important channel through which adviser connectedness affects student placement.

We defined favoritism vis-a-vis the student's adviser as hiring of her student by her coauthors' university, i.e. the student is placed in the adviser's coauthors' department. Bian et al. (2016) study the extent of favoritism with senior hiring in the German academic market. They define favoritism as hirings of senior researchers that were colleagues with individuals working in the hiring department. We assess the importance of this channel by an estimation of the minimum social distance between an adviser and faculty members of the university where her student got placed. Social distance is defined as number of intermediate nodes between any two nodes in the network, given that a path exists between them. For example, if the student is placed in the adviser's coauthors' department, the corresponding social distance would be 1. On the other hand if a student is placed at a department where her adviser has no co-authors but one of her adviser's co-authors has a co-author, the social distance is 2. Figure 4.1 indicates that out of the 519 advisers for whom we can connect adviser and student's placement faculty members (through some path of co-authors), only 11 times did the student go to her adviser's coauthor's university. This further strengthens our assertion that indirect connections matter a lot for job placement (over and above direct connections). In a further 21 cases, the minimum social distance between the adviser and the nearest placement faculty member is 2. The mean social distance between an adviser and her student's placement is about 7. That is, on average there are 6 researchers between the adviser and her student's placement faculty. The number is exactly the same when including the social distances of committee members (which we know for 109 PhD students). Given the high average social distance between the adviser and the student's placement, we conclude that there is little scope for favoritism.[29]

[29] The finding of a relatively high social distance is also interesting in light of the findings of Baruffaldi et al. (2016). The authors relate a PhD student's productivity to where she obtained the previous academic degree. They find PhD students trained at the affiliations of the new supervisor's coauthors are most productive, i.e. where the social distance is non-zero, but small.

Figure 4.1: Histogram of minimum social distance to placement faculty.



Notes: Histogram shows social distance between a student's adviser and the nearest member of the placement faculty. Social distance is the number of nodes on a path between nodes and is measured in the coauthor network of the year of placement.

It could be argued that adviser connectedness helps student placement if some departments are afraid of refusing students of advisers who are influential in their field. However, we believe that by including controls for adviser's productivity, age, gender and affiliation, we control for this effect.

Table 4.8: Results of a logistic regression for placement probability in adviser distance sample.

	Student placement	
	(1)	(2)
Increase in dist after death	-0.363^{*}	-0.361^{*}
	$p = 0.061$	$p = 0.067$
Co-author distance before death	-0.224^{***}	-0.200^{***}
	$p = 0.000$	$p = 0.000$
Euclidean Index	0.000639^{***}	0.000535^{***}
	$p = 0.006$	$p = 0.000$
Experience	0.0152	0.0354
	$p = 0.656$	$p = 0.265$
Experience square	-0.000427	-0.000922
	$p = 0.603$	$p = 0.254$
Placement Rank	-0.00499^{***}	-0.00517^{***}
	$p = 0.000$	$p = 0.000$
Constant	-4.480^{***}	-5.053^{***}
	$p = 0.000$	$p = 0.000$
Adviser fixed effect	*Yes*	*No*
Year fixed effect	*Yes*	*Yes*
PhD School fixed effect	*No*	*Yes*
N	992,079	964,309

Notes: ***, ** and * indicate statistical significance to the 1, 5 and 10 percent level, respectively. All variables refer to dyads between adviser a and placement k in year t, given that k is listed in our version of the Tilburg University Economics ranking and a list of faculty members is available. *Student Placement* equals 1 if a student of adviser a was placed at university k in year t. *Increase in social dist. after death* is the increase in social distance in the co-author network between a and the nearest faculty member of k after a scientist died in year $t-1$. *Social dist. before death* is the social distance in the co-author network between a and the nearest faculty member of k before the exogenous removal of deceased authors. *Placement Rank* is the rank according to our version of the Tilburg University Economics ranking of university k in year t. *Euclidean Index* is the adviser's Euclidean index of citations in year t. *Experience* is the number of years between an adviser's first publication and year t. *Experience*^{2} is its square. *Male adv.* equals 1 if the adviser is estimated to not be female. Only paths between known advisers of students graduating from North-American universities and faculty members of departments from the Hasselback faculty roosters considered.

4.7 Discussion and Conclusion

We show that students receive better placement outcomes when their adviser is better connected in the Economics co-author network. We provide supportive evidence to argue that this could be because more central advisers are better positioned to disseminate information in the network, which ultimately decreases information asymmetry regarding the match quality of her student with a prospective university. Our research is relevant for understanding the placement of graduate students. Since initial placement matters a lot for an Economist's career (Oyer, 2006), the effort dedicated to understanding it, can hardly be overstated.

Furthermore, our result that the connectedness of the adviser matters for the placement of Economics graduates has insights into possible results in the general labor market. Hitherto, several papers have documented that referrals and job opening information from currently employed workers matters for job seeking individuals. However, we demonstrate that not all connections are equal - more connected workers[30] could be more important for job seekers. We also demonstrate through our study that indirect connections could be an important determinant of job market outcome. Finally, due to the special characteristics of the Economics Job Market, one of which is that there is no information asymmetry regarding job openings, we are able to provide some evidence to argue that social networks serve as a conduit of information regarding an applicant's quality.

Further avenues for research include the quality of a job match. Ultimately, the Economics job market is not necessarily about matching the student with the highest ranked department, but to improve the match between the student and the department (Smeets et al., 2006). It would be interesting to see how students matched after recommendations/calls from the adviser fare in the academic world. A good measure of match quality would be if the student gets tenure at the university which first hires the student. Another extension of our work would be to study how adviser connections in industry could affect non-academic placements.

[30] This could be in the network of all jobs held (where two workers are connected if they have ever worked in the same firm at the same time), or in the networks of colleagues, friends and acquaintances.

Appendix

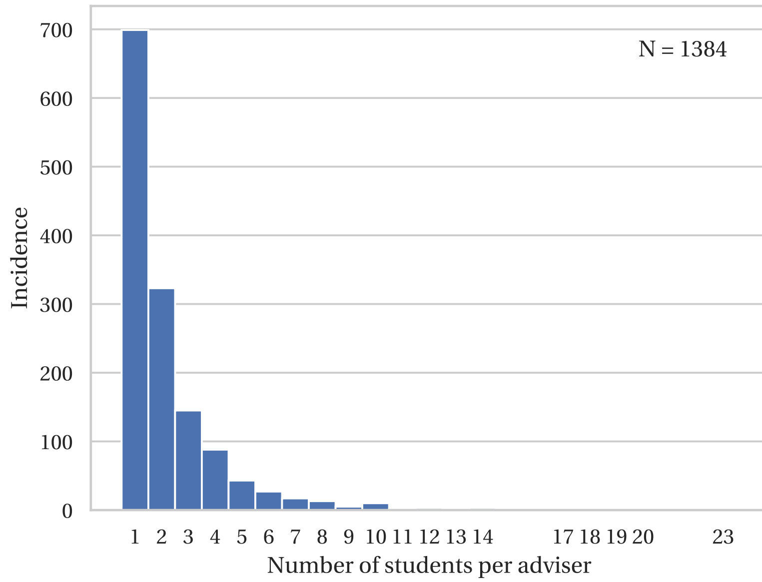
4.A Additional tables and figures

Table 4.A.1: Advisers with most PhD students, 2000-2004.

	Name	Students	University	Citations	Euclid	Seniority
1	Daron Acemoglu	23	Massachusetts Institute of Technology	1258.0	138.40	10
2	Andrei Shleifer	20	Harvard University	11257.0	715.10	17
3	Roger R. Betancourt	19	University of Maryland	430.0	30.84	32
4	Peter C.B. Phillips	18	Yale University	21933.0	1675.89	30
5	John Y. Campbell	17	Harvard University	5457.0	312.60	17
5	Lawrence F. Katz	17	Harvard University	6287.0	606.40	22
7	Arnold C. Harberger	14	University of California, Los Angeles	271.0	33.02	47
7	Olivier Jean Blanchard	14	Massachusetts Institute of Technology	7003.0	578.85	24
7	Ronald Andrew Ratti	14	University of Missouri	103.0	9.70	26
10	Abhijit V. Banerjee	13	Massachusetts Institute of Technology	2248.0	320.74	12
11	Dominick Salvatore	12	Fordham University	612.0	41.04	31
11	John C. Haltiwanger	12	University of Maryland	1125.0	80.04	21
11	Ricardo J. Caballero	12	Massachusetts Institute of Technology	1535.0	146.14	14
14	George W. Evans	11	University of Oregon	1327.0	85.59	19
15	Barry J. Eichengreen	10	University of California, Berkeley	2734.0	128.16	22
15	Gary S. Becker	10	University of Chicago	3478.0	295.30	18
15	James M. Poterba	10	Massachusetts Institute of Technology	5862.0	481.83	20
15	Joshua D. Angrist	10	Massachusetts Institute of Technology	1533.0	231.20	13
15	Larry A. Sjaastad	10	Texas A&M University & University of Chicago	152.0	18.71	43
15	Mohsen Bahmani-Oskooee	10	University of Wisconsin, Milwaukee	1465.0	85.34	19
15	Robert A. Moffitt	10	Johns Hopkins University	2121.0	128.23	26
15	Stephen J. Turnovsky	10	University of Washington	4345.0	143.95	35
15	Thomas D. Willett	10	Claremont Graduate University	1642.0	73.44	35
15	Thomas J. Holmes	10	University of Minnesota	95.0	21.84	13

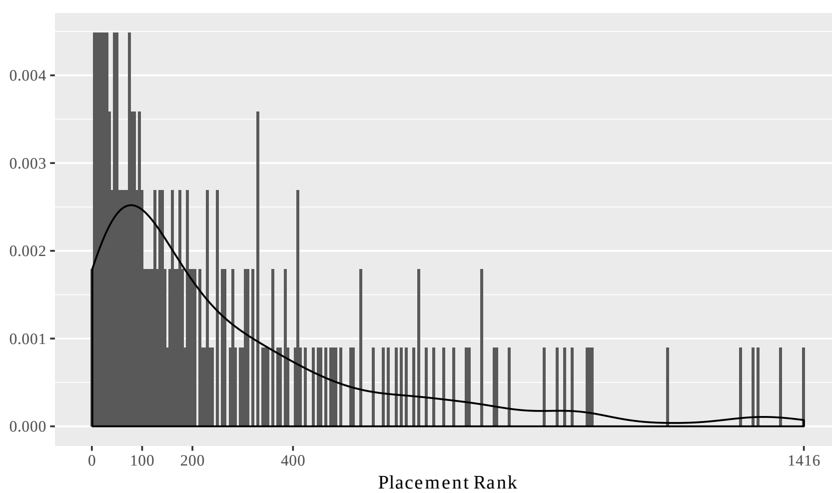
Notes: Table lists PhD advisers by number of PhD students that graduated at North American Economics departments in the academic years 2000/2001, 2001/2002, 2002/2003 and 2003/2004. *Students* is the number of students that graduated with this adviser and that have a Scopus profile. *Citations* is the number of citations to that author. *Publications* is the number of publications of that author. *Seniority* is the number of years since the first publication. All information originate from Scopus and were obtained in March 2017. Only advisers with Scopus profile considered. Co-advised students count as full supervised students.

Figure 4.A.1: Histogram showing the number of students per adviser (academic years 2000/01-2003/04).



Notes: Histogram shows the number of advisers (y axis) with a given number of students (x axis). Only students with known adviser from North-American universities that graduated in the academic years 2000/2001, 2001/2002, 2002/2003, and 2003/2004 considered.

Figure 4.A.2: Distribution of placement rank of initial placement.



Notes: Graph shows the distribution of the initial placement of students in our dataset. Only students with known adviser from North-American universities that graduated in the academic years 2000/2001, 2001/2002, 2002/2003, and 2003/2004 considered, whose initial placement is ranked in the our version of the Tilburg University Economics ranking.

Table 4.A.2: Summary statistics for all continuous variables in the adviser coauthor centrality sample.

Placement Rank		0.21	0.17	0.07	0.52	−0.41	−0.04
Adv. Eigenvector rank	0.23		0.70	0.07	0.19	−0.12	−0.02
Adv. neigh. mean Eigenvector rank	0.18	0.77		0.07	0.20	0.00	0.13
Tightness	0.08	0.06	0.07		−0.09	0.01	−0.01
School Rank	0.28	0.15	0.12	−0.03		−0.51	0.02
Adv. Euclidean Index	−0.20	−0.13	−0.02	−0.14	−0.18		0.34
Adv. Experience	−0.05	−0.03	0.12	−0.01	0.07	0.22	

Notes: Upper triangular depicts Spearman correlation coefficients while lower correlation reports Pearson correlation coefficients. *Placement Rank* is the rank according to our version of the Tilburg University Economics ranking of a student's placement in the year of the placement. *Adv. neigh. mean Eigenvector rank* is the mean Eigenvector centrality rank of all coauthors of an adviser in the weighted coauthor network corresponding to the year of the placement. *Market tightness* is the field-specific number of students who graduated in a year divided by the in this field number of AEA-reported job openings for that year (equation (4.1)). *School Rank* is the rank according to our version of the Tilburg University Economics ranking of the PhD-awarding university in the year the student finished. *Euclidean Index* is the adviser's Euclidean index of citations in the year the student graduated. *Experience* is the number of years between an adviser's first publication and the year in which the student graduated. *Experience*^{2} is its square. Only students considered that were placed at a ranked institution, whose adviser is in the network's giant component, and whose adviser has students in different years for which above two conditions hold.

Table 4.A.3: List of deceased faculty members in the dataset.

Name	Date of death	Name	Date of death
Dalton, George	1999, Aug 23	Vilasuso, Jon R.	2002, Apr 27
Liu, Jung-Chao	1999, Aug 31	Bowman, Mary Jean	2002, Jun 04
Griliches, Zvi	1999, Nov 04	Smith, Bruce D.	2002, Jul 09
Gapinski, James H.	2000, Jan 01	Ansoff, H. Igor	2002, Jul 14
Johnson, Byron L.	2000, Jan 06	Dornbusch, Rüdiger	2002, Jul 25
Heyne, Paul	2000, Mar 09	Ando, Albert	2002, Sep 19
Miller, Merton H.	2000, Jun 03	Gabriel, Stuart A.	2002, Oct 15
Lillard, Lee A.	2000, Dec 02	Sertel, Murat R.	2003, Jan 25
Elliott, John E.	2001, Jan 01	Johnson, D. Gale	2003, Apr 13
Cameron, Rondo	2001, Jan 01	Berger, Mark C.	2003, Apr 30
Cookingham, Mary E.	2001, Mar 12	Kain, John F.	2003, Aug 03
Rosen, Sherwin	2001, Mar 17	Modigliani, Franco	2003, Sep 25
Moses, Ronald	2001, Jun 20	Hsu, Robert	2004, Jan 18
Straub, LaVonne	2002, Jan 24	Lee, Winson	2004, Mar 01
Rosenthal, Robert W.	2002, Feb 07	Laffont, Jean Jacques	2004, May 01

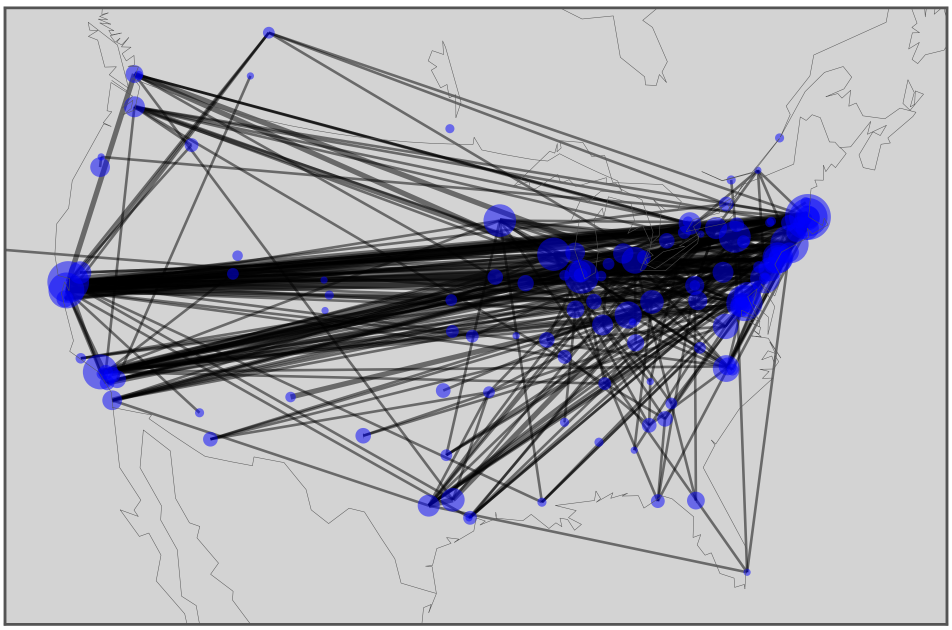
Notes: Table lists 30 authors who passed away between summer 1999 and summer 2004 while serving on the faculty of universities as listed in the Hasselback lists.

Table 4.A.4: Correlations for placement probability in the adviser distance sample.

Student placement		1.00	0.00	0.02	0.00	0.00
Increase in social dist. after death	0.00		−0.03	0.00	1.00	0.00
Social dist. before death	−0.03	−0.03		−0.03	−0.01	−0.03
Placement Rank	0.01	−0.01	−0.14		−0.02	−0.01
Euclidean Index	0.00	0.00	−0.05	0.28		1.00
Experience	−0.03	0.01	0.40	0.01	0.02	

Notes: Upper triangular depicts Spearman correlation coefficients while lower correlation reports Pearson correlation coefficients. All variables refer to time-variant dyads between adviser a and placement k in year t, given that k appears in the our ranking and a list of faculty members is available. *Student Placement* equals 1 if a student of adviser a was placed at university k in year t. *Increase in social dist. after death* is the increase in social distance in the co-author network between a and the nearest faculty member of k after scientist died in year $t-1$. *Social dist. before death* is the social distance in the co-author network between a and the nearest faculty member of k before the distance changed induced by the removal of deceased scientists. *Placement Rank* is the rank according to our version of the Tilburg University Economics ranking of university k in year t. *Euclidean Index* is the adviser's Euclidean index of citations in year t. *Experience* is the number of years between an adviser's first publication and year t. *Male adv.* equals 1 if the adviser is estimated to not be female. Only paths between known advisers of students graduating from North-American universities and identified faculty members of departments listed in our version of the Tilburg University Economics ranking and available Hasselback faculty roosters considered.

Figure 4.A.3: Hiring network of North American universities 2000/2001-2003/2004.



Notes: Map shows hiring network for North American universities for the academic years 2000-2001, 2001-2002, and 2002-2003. Every node represents a university from which at least one student graduated that was subsequently hired by another university on the map, which is indicated by the links (Nodes representing Hawaiian universities are omitted). Nodes are sized according to how many students graduated from that university. Network is calculated from the placement of 451 students going from/to 132 universities.

Chapter 5

Conclusion

As probably all dissertations in Economics, this thesis' title page was followed by acknowledgements. I am not alone doing so. Acknowledgements are customary and widespread in Economics and beyond. In it, the reader finds names of other scientists, of research assistants, of industry professionals, of sources of funding, of universities where seminars have been given and of conferences the current research has been presented. Scientists use acknowledgements to express gratitude for intellectual and other help during various stages of their research projects, without transferring intellectual property rights, which remain solely with the author(s) (Cronin, 1995; Laband and Tollison, 2000). This thesis uses these acknowledgements as primary source of data to create a social network connecting collaborators, which allows us to study informal collaboration, and to investigate the corresponding social networks.

Because of its widespread use, it surprises how little research is available on the importance of informal collaboration and how rarely acknowledgements are used as proxy for intellectual collaboration. Going forward, I therefore plan to examine how the financial economics academe evolves from a network perspective, using the data on informal collaboration. Two groups of questions emerge, of which the first group aims at the network structure: How has the network changed since the financial crisis? Is the degree of connectedness prone to favor collusion? Can we identify schools of thought and if yes, how does this matter for the progress of science? The financial crisis and the high quality of my data furthermore allows to study Kuhn's statement that the scientific progress is due to paradigm shifts and the structure of the scientific community. For example, Kuhn hypothesizes that only competition between segments of the scientific community results in the rejection of one previously accepted theory or in the adoption of another.

A second group of questions looks at individual network behavior and group formation: Which researchers do authors choose to work with? Should individual networks of researchers be taken into account in hiring decisions? Do authors and informal collaborators complement or enhance each other e.g. by skill set? In particular, is skill match in groups dependent on whether its members are generalists or specialists? How does group formation affect individual careers? Do male and female researchers network differently? A particular concern arises from the finding presented in chapter 2.3.6 that females are less central and less often acknowledged regardless of academic prolificness and experience. This above all warrants more research.

Bibliography

Abadie, A., Athey, S., Imbens, G. and Wooldridge, J. (2017), ‘When Should You Adjust Standard Errors for Clustering?’, *National Bureau of Economic Research, Working Paper*.

Ad Hoc Committee on the Job Market (2011), ‘Report of Ad Hoc Committee on the Job Market’, *American Economic Review* **101**(3), 744–746.

Adams, J. D. (1990), ‘Fundamental Stocks of Knowledge and Productivity Growth’, *Journal of Political Economy* **98**(4), 673–702.

Alatas, V., Banerjee, A., Chandrasekhar, A. G., Hanna, R. and Olken, B. A. (2016), ‘Network Structure and the Aggregation of Information: Theory and Evidence from Indonesia’, *The American Economic Review* **106**(7), 1663–1704.

Athey, S., Katz, L. F., Krueger, A. B., Levitt, S. and Poterba, J. (2016), ‘What Does Performance in Graduate School Predict? Graduate Economics Education and Student Outcomes’, *The American Economic Review* **97**(2), 512–518.

Azoulay, P., Fons-Rosen, C. and Zivin Graff, J. S. (2015), ‘Does Science Advance One Funeral at a Time?’, *National Bureau of Economic Research, Working Paper Series*.

Azoulay, P., Graff Zivin, J. S. and Wang, J. (2010), ‘Superstar Extinction’, *The Quarterly Journal of Economics* **125**(2), 549–589.

Ballester, C., Calvó-Armengol, A. and Zenou, Y. (2006), ‘Who’s Who in Networks. Wanted: The Key Player’, *Econometrica* **74**(5), 1403–1417.

Banerjee, A., Chandrasekhar, A. G., Duflo, E. and Jackson, M. O. (2013), 'The Diffusion of Microfinance', *Science* **341**(6144), 1236498–1236498.

Banerjee, A., Chandrasekhar, A. G., Duflo, E. and Jackson, M. O. (2014), 'Gossip: Identifying Central Individuals in a Social Network', *NBER Working Paper 20422*.

Baruffaldi, S. H., Visentin, F. and Conti, A. (2016), 'The productivity of science & engineering PhD students hired from supervisors' networks', *Research Policy* **45**(4), 785–796.

Bavelas, A. (1950), 'Communication Patterns in Task-Oriented Groups', *The Journal of the Acoustical Society of America* **22**(6), 725–730.

Bayer, P., Ross, S. L. and Topa, G. (2008), 'Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes', *Journal of Political Economy* **116**(6), 1150–1196.

Beaver, D. d. and Rosen, R. (1979), 'Studies in scientific collaboration Part II. Scientific coauthorship, research productivity and visibility in the French scientific elite, 1799–1830', *Scientometrics* **1**(2), 133–149.

Bian, B., Haselman, R. and Vig, V. (2016), 'Incentives , Hiring and Productivity: Evidence from Academia'.

Bonacich, P. (1987), 'Power and Centrality: A Family of Measures', *American Journal of Sociology* **92**(5), 1170–1182.

Boorman, S. A. (1975), 'A Combinatiorial Optimization Model for Transmission of Job Information through contact networks', *The Bell Journal of Economics* **6**(1), 216–249.

Borokhovich, K. A., Bricker, R. J. and Simkins, B. J. (2000), 'An Analysis of Finance Journal Impact Factors', *The Journal of Finance* **55**(3), 1457–1469.

Bramoullé, Y., Djebbari, H. and Fortin, B. (2009), 'Identification of peer effects through social networks', *Journal of Econometrics* **150**(1), 41–55.

Bramoullé, Y. and Kranton, R. (2007), 'Public goods in networks', *Journal of Economic Theory* **135**(1), 478–494.

Bramoullé, Y., Kranton, R. and Amours, M. D. (2014), 'Strategic Interaction and Networks', *The American Economic Review* **104**(May), 898–930.

Brogaard, J., Engelberg, J. and Parsons, C. A. (2014), 'Networks and productivity: Causal evidence from editor rotations', *Journal of Financial Economics* **111**(1), 251–270.

Brown, L. D. (2005), 'The importance of circulating and presenting manuscripts: Evidence from the accounting literature', *Accounting Review* **80**(1), 55–83.

Burks, S. V., Cowgill, B., Hoffman, M. and Housman, M. (2017), 'The Value of Hiring through Employee Referral', *The Quarterly Journal of Economics* **130**(2), 805–839.

Calvó-Armengol, A. (2004), 'Job contact networks', *Journal of Economic Theory* **115**(1), 191–206.

Calvó-Armengol, A., Patacchini, E. and Zenou, Y. (2009), 'Peer effects and social networks in education', *The Review of Economic Studies* **76**(4), 1239–1267.

Calvó-Armengol, A. and Zenou, Y. (2005), 'Job matching, social network and word-of-mouth communication', *Journal of Urban Economics* **57**(3), 500–522.

Card, D. and DellaVigna, S. (2013), 'Nine Facts about Top Journals in Economics', *Journal of Economic Literature* **51**(1), 144–161.

Chari, A. and Goldsmith-Pinkham, P. (2017), 'Gender Representation in Economics Across Topics and Time: Evidence from the NBER Summer Institute', *NBER Working Paper 23953*.

Cohen, W. M. and Levinthal, D. A. (1989), 'Innovation and Learning: The Two Faces of R & D', *The Economic Journal* **99**(397), 569–596.

Cohen, W. M., Nelson, R. R. and Walsh, J. P. (2002), 'Links and Impacts: The Influence of Public Research on Industrial R&D', *Management Science* **48**(1), 1–23.

Colander, D. (1989), 'Research on the Economics Profession', *Journal of Economic Perspectives* **3**(4), 137–148.

Colander, D. (1997), Surviving as a Slightly Out of Sync Economist, *in* S. G. Medema and W. J. Samuels, eds, 'Foundations of Research in Economics: How Do Economists Do Economics?', Edward Elgar, Cheltenham, UK and Northampton, MA.

Coles, P., Cawley, J., Levine, P. B., Niederle, M., Roth, A. E. and Siegfried, J. J. (2010), 'The Job Market for New Economists: A Market Design Perspective', *Journal of Economic Perspectives* **24**(4), 187–206.

Colussi, T. (2017), 'Social Ties in Academia: A Friend Is a Treasure', *The Review of Economics and Statistics* **98**(2).

Combes, P.-p. and Linnemer, L. (2010), 'Inferring missing citations: a quantitative multi-criteria ranking of all jornals in economics', *HAL, Working Paper*.

Conley, T. G. and Udry, C. R. (2010), 'Learning about a New Technology: Pineapple in Ghana', *The American Economic Review* **100**(1), 35–69.

Conti, A. and Visentin, F. (2015), 'A revealed preference analysis of PhD students' choices over employment outcomes', *Research Policy* **44**(10), 1931–1947.

Coupé, T. (2004), 'What Do We Know about Ourselves? on the Economics of Economics', *Kyklos* **57**(2), 197–215.

Cronin, B. (1995), *The Scholar's Courtesy: The Role of Acknowledgment in the Primary Communication Process*, Taylor Graham, Los Angeles, CA.

Cruz, C., Labonne, J. and Querubin, P. (2017), 'Politician Family Networks and Electoral Outcomes: Evidence from the Philippines', *American Economic Review* **107**(10), 1–34.

Dahlander, L. and McFarland, D. A. (2013), 'Ties That Last: Tie Formation and Persistence in Research Collaborations over Time', *Administrative Science Quarterly* **58**(1).

Debreu, G. and Herstein, I. N. (1953), 'Nonnegative Square Matrices', *Econometrica* **21**(4), 597–607.

Ductor, L. (2015), 'Does Co-authorship Lead to Higher Academic Productivity?', *Oxford Bulletin of Economics and Statistics* **77**(3), 385–407.

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

Ductor, L., Goyal, S. and Prummer, A. (2018), 'Gender & Collaboration', *Cambridge-INET Institute, Working Paper Series*.

Dustmann, C., Glitz, A., Schönberg, U. and Brücker, H. (2016), 'Referral-based Job Search Networks', *The Review of Economic Studies* **83**(2), 514–546.

Eagly, R. V. (1975), 'Economics Journals as a Communications Network', *Journal of Economic Literature* **13**(3), 878–888.

Elliott, M. and Golub, B. (n.d.), 'A Network Approach to Public Goods', *Journal of Political Economy* .

Ellison, G. (2002), 'The Slowdown of the Economics Publishing Process', *Journal of Political Economy* **110**(5), 947–993.

Fourcade, M., Ollion, E. and Algan, Y. (2015), 'The Superiority of Economists', *Journal of Economic Perspectives* **29**(1), 89–114.

Fracassi, C. and Tate, G. (2012), 'External Networking and Internal Firm Governance', *The Journal of Finance* **67**(1), 153–194.

Freeman, L. C. (1978), 'Centrality in social networks conceptual clarification', *Social Networks* **1**(3), 215–239.

Freeman, R. B. and Huang, W. (2015), 'Collaborating with People Like Me: Ethnic Coauthorship within the United States', *Journal of Labor Economics* **33**(S1), S289–S318.

Friedkin, N. E. (2015), 'Theoretical Measures ' Foundations for Centrality', *Scientometrics* **96**(6), 1478–1504.

Gallet, C. A., List, J. A. and Orazem, P. F. (2005), 'Cyclicality and the Labor Market for Economists', *Southern Economic Journal* **72**(2), 284.

Gaulé, P. and Piacentini, M. (2017), 'An Advisor Like Me? Advisor Gender and Post-graduate Careers in Science', *IZA, Discussion Papers Series*.

Girotra, K., Terwiesch, C. and Ulrich, K. T. (2010), 'Idea Generation and the Quality of the Best Idea', *Management Science* **56**(4), 591–605.

Goldsmith-Pinkham, P. and Imbens, G. W. (2013), 'Social Networks and the Identification of Peer Effects', *Journal of Business & Economic Statistics* **31**(3), 253–264.

Goyal, S. and Moraga-González, J. L. (2001), 'R&D Networks', *The RAND Journal of Economics* **32**(4), 686–707.

Goyal, S., van der Leij, M. J. and Moraga-González, J. L. (2006), 'Economics: An Emerging Small World', *Journal of Political Economy* **114**(2), 403–412.

Graham, B. S. (2015), 'Methods of Identification in Social Networks', *Annual Review of Economics* **7**(1), 465–485.

Granovetter, M. S. (1973), 'The Strength of Weak Ties', *American Journal of Education* **78**(6), 1360–1380.

Green, R., O'Hara, M. and Schwert, G. W. (2002), 'Joint Editorial: Advice for Authors', *The Journal of Finance* **57**(2), 1031–1032.

Haeussler, C. and Sauermann, H. (2016), 'The Division of Labor in Teams: A Conceptual Framework and Application to Collaborations in Science', *NBER Working Paper 22241*.

Hamermesh, D. S. (1992), 'The Young Economist's Guide to Professional Etiquette', *Journal of Economic Perspectives* **6**(1), 169–179.

Hensvik, L. and Skans, O. N. (2016), 'Social Networks, Employee Selection, and Labor Market Outcomes', *Journal of Labor Economics* **34**(4), 825–867.

Hinshaw, C. E. (2000), 'Report of the Director Job Openings for Economists', *American Economic Review* **89**(2), 484–485.

Hinshaw, C. E. (2001), 'Report of the Director Job Openings for Economists', *American Economic Review* **89**(2), 484–485.

Hinshaw, C. E. (2002), 'Report of the Director Job Openings for Economists', *American Economic Review* **89**(2), 484–485.

Hinshaw, C. E. (2003), 'Report of the Director Job Openings for Economists', *American Economic Review* **89**(2), 484–485.

Hinshaw, C. E. (2004), 'Report of the Director Job Openings for Economists', *American Economic Review* **89**(2), 484–485.

Hojman, D. A. and Szeidl, A. (2008), 'Core and periphery in networks', *Journal of Economic Theory* **139**(1), 295–309.

Hollis, A. (2001), 'Co-authorship and the output of academic economists', *Labour Economics* **8**(4), 503–530.

Hsieh, C.-S., König, M. D., Liu, X. and Zimmermann, C. (2018), 'Superstar Economists: Coauthorship networks and research output', *CEPR Discussion Paper*.

Hudson, J. (1996), 'Trends in Multi-Authored Papers in Economics', *Journal of Economic Perspectives* **10**(3), 153–158.

Ioannides, Y. M. (2012), *From Neighborhoods to Nations: The Economics of Social Interactions*, Princeton University Press, Princeton, New Jersey.

Jackson, M. O. (2014), 'Networks in the Understanding of Economic Behaviors', *Journal of Economic Perspectives* **28**(4), 3–22.

Jackson, M. O., Rogers, B. W. and Zenou, Y. (2017), 'The Economic Consequences of Social-Network Structure', *Journal of Economic Literature* **55**(1), 49–95.

Jackson, M. O. and Wolinsky, A. (1996), 'A Strategic Model of Social and Economic Networks', *Journal of Economic Theory* **71**(1), 44–74.

Jackson, M. O. and Zenou, Y. (2015), Games on Networks, *in* H. P. Young and S. Zamir, eds, 'Handbook of Game Theory', Elsevier B.V., chapter 3, pp. 95–163.

Jaffe, A. B. (1989), 'Real Effects of Academic Research', *The American Economic Review* **79**(5), 957–970.

Jones, B. F. (2009), 'The Burden of Knowledge and the "Death of the Renaissance Man": Is Innovation Getting Harder?', *The Review of Economic Studies* **76**, 283–317.

König, M. D., Liu, X. and Zenou, Y. (2014), 'R&D Networks: Theory, Empirics and Policy Implications', *CEPR, Discussion Paper*.

Kramarz, F. and Skans, O. N. (2014), 'When Strong Ties are Strong: Networks and Youth Labour Market Entry', *The Review of Economic Studies* **81**(3), 1164–1200.

Krueger, A. B. and Wu, S. (2000), 'Forecasting Job Placements of Economics Graduate Students', *The Journal of Economic Education* **31**(1), 81.

Laband, D. N. and Tollison, R. D. (2000), 'Intellectual Collaboration', *Journal of Political Economy* **108**(3), 632–661.

Laband, D. N. and Tollison, R. D. (2003), 'Good colleagues', *Journal of Economic Behavior & Organization* **52**(4), 505–512.

Li, E. Y., Liao, C. H. and Yen, H. R. (2013), 'Co-authorship networks and research impact: A social capital perspective', *Research Policy* **42**(9), 1515–1530.

Mansfield, E. (1991), 'Academic research and industrial innovation', *Research Policy* **20**(1), 1–12.

Manski, C. F. (1993), 'Identification of Endogenous Social Effects: The Reflection Problem', *The Review of Economic Studies* **60**(60), 531–542.

Mansour, H., Rees, D. I., Rintala, B. M. and Wozny, N. N. (2018), 'The Effects of Professor Gender on the Post-Graduation Outcomes of Female Students', *IZA Discussion Paper Series*.

McDowell, J. M. and Melvin, M. (1983), 'The Determinants of Co-Authorship: An Analysis of the Economics Literature', *The Review of Economics and Statistics* **65**(1), 155.

Medoff, M. H. (2003), 'Collaboration and the quality of economics research', *Labour Economics* **10**(5), 597–608.

Mohnen, M. (2016), 'Stars and Brokers : Peer Effects among Medical Scientists'.

Montgomery, J. D. (1991), 'Social Networks and Labor-Market Outcomes: Toward an Economic Analysis', *The American Economic Review* **81**(5), 1408–1418.

Mullahy, J. (1986), 'Specification and testing of some modified count data models', *Journal of Econometrics* **33**(3), 341–365.

Oettl, A. (2012), 'Reconceptualizing Stars: Scientist Helpfulness and Peer Performance', *Management Science* **58**(6), 1122–1140.

Oyer, P. (2006), 'Initial Labor Market Conditions and Long-Term Outcomes for Economists', *Journal of Economic Perspectives* **20**(3), 143–160.

Perry, M. and Reny, P. J. (2016), 'How To Count Citations If You Must', *The American Economic Review* **106**(9), 2722–2741.

Phelps, C., Heidl, R. and Wadhwa, A. (2012), 'Knowledge, Networks, and Knowledge Networks: A Review and Research Agenda', *Journal of Management* **38**(4), 1115–1166.

Ponomariov, B. and Boardman, C. (2016), 'What is co-authorship?', *Scientometrics* **109**(3), 1939–1963.

Singh, J. (2005), 'Collaborative Networks as Determinants of Knowledge Diffusion Patterns', *Management Science* **51**(5), 756–770.

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

Smeets, V., Warzynski, F. and Coupé, T. (2006), 'Does the Academic Labor Market Initially Allocate New Graduates Efficiently?', *Journal of Economic Perspectives* **20**(3), 161–172.

Stephan, P. E. (2012), *How economics shapes science*, Harvard University Press, Cambridge, MA.

Storper, M. and Venables, A. J. (2004), 'Buzz: face-to-face contact and the urban economy', *Journal of Economic Geography* **4**(4), 351–370.

Thomson, W. (1999), 'The Young Person's Guide to Writing Economic Theory', *Journal of Economic Literature* **37**(1), 157–183.

Waldinger, F. (2012), 'Peer Effects in Science: Evidence from the Dismissal of Scientists in Nazi Germany', *The Review of Economic Studies* **97**(2), 838–861.

Watts, D. J. (1999), 'Networks, Dynamics, and the Small-World Phenomenon', *American Journal of Sociology* **105**(2), 493–527.

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

Wu, A. H. (2018), 'Gendered Language on the Economics Job Market Rumors Forum', *American Economic Association Papers and Proceedings* **108**, 175–79.

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