KIEL
WORKING PAPER

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

No. 2182 April 2021

Michael E. Rose, and Co-Pierre Georg

Kiel Institute for the World Economy
ISSN 1862–1155

ABSTRACT

WHAT 5,000 ACKNOWLEDGEMENTS TELL US ABOUT INFORMAL COLLABORATION IN FINANCIAL ECONOMICS*

Michael E. Rose, and Co-Pierre Georg

We present and discuss a novel dataset on informal collaboration in financial economics, manually collected from more than 5,000 acknowledgement sections of published papers. We find that informal collaboration is the norm in financial economics, while generational differences in informal collaboration exist and reciprocity among collaborators prevails. Female researchers appear less often in acknowledgements than comparable male researchers. Information derived from networks of informal collaboration allows us to predict academic impact of both researchers and papers even better than information from co-author networks. Finally, we study the characteristics of the networks using various measures from network theory and characterize what determines a researcher's position in it. The data presented here may help other researchers to shed light on an under-explored topic.

Keywords: intellectual collaboration, acknowledgements, social networks, financial economics

JEL classification: A14; D83; G00; O33

Michael E. Rose (corresponding author)
Kiel Institute for the World Economy
Kiellinie 66
D-24105 Kiel, Germany

Max Planck Institute for Innovation and Competition
Marstallplatz 1,
80539 München, Germany
Email: michael.rose@ip.mpg.de
www.ip.mpg.de

Co-Pierre Georg
University of Cape Town
Lover'sWalk, Rondebosch
7700, Cape Town, South Africa
Email: co-pierre.georg@uct.ac.za
www.uct.ac.za

* This previously circulated as "Mirror, Mirror On The Wall, Who is the Most Central of Them All?". It is derived from chapter 2 of Michael's dissertation at the University of Cape Town. Special thanks to the dissertation committee members Marco van der Leij, Matthew Elliott, and Lorenzo Ductor. For very helpful suggestions and comments that led to a substanial improvement of the paper we thank Adam Jaffe (the acting editor) and two referees. We are furthermore grateful to Stefano Baruffaldi, Krishna Dasaratha, Falko Fecht, Fabian Gaessler, Michael Gofman, Nir Hak, Han Kim, Alex Oettl, Daniel Opolot, and Suraj Shekhar. The paper als benefited from the helpful audience in seminars at the Kiel Institute for the World Economy, George Mason University, the University of Amsterdam, the Free University of Amsterdam, and University of Fribourg, as well as at the 2017 Southern African Finance Association meeting. Nurzhan Sapargali provided able research assistance and Myriam Rion provided superb editorial assistance.

1 Introduction

Collaboration is prevalent in academic research and research conducted in teams of co-authors tends to be more influential than single-authored work (Beaver and Rosen, 1978; Wuchty et al., 2007). Yet, collaboration through co-authorship is only one form of collaboration among academics (Cronin, 1995; Katz and Martin, 1997). Laband and Tollison (2000) highlight that informal intellectual collaboration, e.g. commenting on a paper or discussing it at a conference or seminar, are common in academic research in economics.[1] Still, little is known about the extent, causes and patterns of informal collaboration, least on the impact on authors' productivity or the academic success of research papers. Certainly the difficulty to observe informal collaboration in most settings is one big hurdle. Our paper fills this gap. We present and discuss a novel and unique dataset on informal collaboration created from metadata of published papers in financial economics. We derive the data from acknowledgements to commenters who provided valuable feedback, and to seminars and conferences where papers were presented.

Our dataset covers informal collaboration as encoded in 5,784 acknowledgement sections, with a focus on interactions of 14,787 distinct researchers.[2] We provide the Scopus author profile ID for 81% of all researchers (including all authors), many of which were verified manually. We source the acknowledgements from 6,597 full research papers published in 6 journals between 1997 and 2011. We draw from the so-called top three finance journals (the Journal of Finance, the Journal of Financial Economics, and the Review of Financial Studies) and a set of three journals with lower impact factor but comparable topical focus (the Journal of Financial Intermediation, the Journal of Money, Credit, & Banking, and the Journal of Banking and Finance).

We distinguish between commenters–those who are acknowledged–and authors–those who wrote the papers. Only about 11.3% of researchers in our dataset fall in both categories. A majority of 56% of all researchers are only commenters but not authors (in the dataset). This already shows the relevance of our dataset, since the commenters' contribution to the published papers would not be observable without looking at acknowledgements. Furthermore, we observe a higher extensive and intensive margin of informal collaboration in the top journals, supporting the hypothesis that informal collaboration may be a driver of higher quality research. We also find an increase in informal collaboration over time. This trend is driven by young authors, rather than by a cultural shift. We find that reciprocity is prevalent within the bounds of our dataset. Up to now, the literature only shows the existence of reciprocity for data sharing (Haeussler et al., 2014). However, reciprocity alone does not explain the large amount of informal collaboration.

We find positive but rather weak correlations (around 0.5) between scientific productivity measures and the number of acknowledged contributions. It seems that a very prolific researcher is not necessarily very "helpful"–a term coined by Oettl (2012*a*,*b*) to describe someone who is acknowledged often. Yet, more prolific researchers tend to be acknowledged more often, as do more senior researchers, but at a decreasing rate. We also find gender differences: Over a three-

[1] In a joint editorial, Green et al. (2002, p. 1032) advise authors to *"circulate their papers and give seminars to colleagues to receive constructive criticism before submitting to a journal."*

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

year period, female researchers are acknowledged in 0.4 fewer papers and by 0.6 fewer authors than their male counterparts. This descriptive finding adds to recent debates regarding the role of gender in informal collaboration (Chari and Goldsmith-Pinkham, 2017), collaboration networks (Ductor et al., 2018) and the perception of female researchers (Sarsons, 2017; Hengel, 2020), and promises to be a relevant piece in the gender gap puzzle.

Collaboration, both between co-authors and between authors and commenters, leads to the emergence of social networks. So far the literature focused on networks among co-authors and established a strong link between an author's position in such networks and a her productivity. The question is whether we find a similar relationship for networks of informal collaboration, and whether such networks contain information not captured by co-author networks. Colander (1989, p. 146) nurtures this idea when he concludes that *"[i]n studying the economics profession, one quickly learns the importance of informal networks, contacts and the exchange of ideas. Much if not most of the debate and discussion about economic ideas take place at the pre-working paper, workshop and working paper stages."*

It has several benefits to connect researchers in a social network based on their informal collaboration. First, network analysis allows studying the speed of learning and the diffusion of information (Alatas et al., 2016). This is due to the fact that social networks are a necessary conduit for tacit information. Second, the network topology may also tell us about the (changing) nature of the profession, revealing, for example its inclusiveness. And third, the individual researcher's position in the network might be consequential for her career and her research output. Access to traversing knowledge but also access to one's social capital are underlying causes (Li et al., 2013). Ductor et al. (2014) hypothesize that closeness to more productive researchers begets early access to new ideas which in turn improves academic output. By repeating the forecast exercise of Ductor et al. (2014) with networks of informal collaboration, we show that networks of informal collaboration capture dimensions distinct from co-author networks. The network of informal collaborations may finally prove useful as an explanatory variable when studying, e.g., the determinants of scientific productivity. Moreover, it may serve as an interesting dependent variable in studies exploring the relationship of individual characteristics and social capital.

Networks of informal collaboration link researchers whenever one acknowledges another. Links are weighted and directed bearing the meaning "has given advice x times". We construct 13 such networks for every year between 1999 and 2011, where each network is inferred from publications over the current year and the preceding two years to account for publication lag. We contrast these networks with common co-author networks where links between researchers indicate the number of joint publications in that period. The networks of informal collaboration are much more connected than the co-author networks created from the same set of papers. The networks of informal collaboration exhibit small-world properties throughout the study period. Small-world networks are particularly interesting since they characterize many real-world networks well and have unique information diffusion properties (Watts and Strogatz, 1998; Watts, 1999).

For each researcher, we compute degree, betweenness centrality, and eigenvector centrality in the network of informal collaboration. These centralities measure how "well connected" the researcher is. They differentiate researchers according to their access to information traversing the network, or to their possible influence on peers (Jackson, 2014; Li et al., 2013). None of these centralities correlate strongly with each other. Most importantly, they do not correlate strongly with measures of academic productivity. This underpins the fact that they capture different aspects of academic life and that informal collaboration is not just a proxy for easier-to-observe attributes.

The network analysis yields a number of insights and stylized facts. We show that the inclusion of an authors' centrality in the network of informal collaboration improves the forecast of future productivity even when controlling for recent past productivity and centrality in the network of formal collaboration. We find that papers that receive feedback from well-connected colleagues have a higher academic impact. Including authors' centralities in a model of a paper's likelihood to get published in a top finance journal and its future number of citations also improves the model forecast. We find that authors who are more eigenvector central in the network of informal collaboration—better positioned to exert influence on their peers—are more likely to publish their paper in a leading finance journal and to receive more citations than the average paper in the journal. Authors who, on the other hand, connect disparate communities, are *less* likely to publish their paper in a leading finance journal. Yet, when they publish it, their paper also receives more citations than the average paper published in this journal. Not claiming causality, we provide evidence for the relevance of informal collaboration and access to knowledge in the academic production function.

Our paper and data are most closely linked to the literature on informal collaboration in academia. Cronin (1995) lays the foundations for the study of informal collaboration in academia as evidenced through acknowledgements.[3] Comparing informal collaboration in economics and biology, Laband and Tollison (2000) find that a higher number of commenters is associated with a higher citation count. Our approach, which is to add information about an academic's position in the social network of informal collaboration, considerably extends the work of Laband and Tollison (2000), who neither study the network of informal collaboration, nor the relationship between informal collaboration and authors' productivity. Brown (2005) includes seminar presentations as another form of informal collaboration and finds that the number of acknowledged seminars is more relevant for the citation count than the number of commenters.[4] Oettl (2012*b*) investigates the causal impact of informal collaboration on authors in the field of immunology. Losing helpful co-authors leads to a drop in the quality of a researcher's output by 14%.

Recently scholars developed an interest in a broader view and to look at venues where informal collaboration takes place. Conferences increase the likelihood that two authors collaborate (Campos et al., 2018). But conferences also act as means to disseminate and possibly improve the presented research (de Leon and McQuillin, 2020; Rose et al., 2020).

Informal collaboration does not only matter in academia, but also in other highly creative activ-

[3] Desrochers et al. (2017) summarize the research on acknowledgements which, since the beginning, includes much more than informal collaboration.

[4] To strengthen the external validity of our data, we replicate the main findings of both of these studies in appendix B.2.

ities. The innovation literature has long acknowledged the importance of close social proximity and the spill-over of tacit knowledge (Buenstorf and Schacht, 2013; Orazbayev, 2017; Andrews, 2019). Informal collaboration enables and facilitates the flow of knowledge, but also helps in the production process—even during the ideation phase (Hasan and Koning, 2019).

We proceed by defining and explaining our data and variables in the next section. The following three sections highlight stylized facts regarding three units of analysis: The acknowledgement section itself, bilateral informal collaboration (the link between commenter and author), and the network of informal collaboration. The last section provides an outlook on a number of interesting research questions our work facilitates.

2 Data and Variables

The main concern regarding acknowledgements as a data source is that authors could use them strategically, e.g., to influence editors, referees, or readers. Authors might want to influence the editors' choice of referees, or increase the referees' and readers' perception of the quality of a paper. While strategic acknowledging is discouraged and can be costly (Hamermesh, 1992), it remains the main threat to measure informal collaboration with acknowledgements data. In appendix C, we discuss various stylized facts that speak against authors' strategic use of acknowledgements.

Sample Selection We measure informal collaboration between authors and commenters through acknowledgments from published papers in financial economics. For two reasons, the topical focus of our paper is the finance subfield of economics. First, informal collaboration is prevalent in economics as opposed to many if not all natural sciences (Laband and Tollison, 2000). And second, finance is a large but homogeneous subfield within economics which alleviates potential concerns about differing social norms in different subfields.

We include the following six journals: 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 JF, RFS and JFE are commonly regarded as the top journals in Financial Economics.[5] The other three journals publish a similar number of papers. We deliberately do not use journals with impact factors close to the top 3 finance journals, such as the Review of Finance and the Journal of Financial and Quantitative Analysis. This enables us to clearly distinguish collaboration in top journals. An analysis of JEL codes confirms the high topical focus.[6]

We include all documents labeled by Scopus as either article, short article, review, conference proceeding, or note. However, we exclude presidential addresses published in the JF, and comments and replies to comments published in any journal. In total we use 5,784 papers.

[5] Borokhovich et al. (2000) and the annual reports in the JF refer to these journals as top journals, too.

[6] 88% of the papers for which we observe the JEL code belong to the general category G (Financial Economics). Additional 8% list E (Macroeconomics and Monetary Economics), but not G. The number of papers with JEL code is 4,867.

Informal Collaboration From each paper's acknowledgement section (which appears often as footnote to the title), we collect the names and number of seminars and conferences, and, the names of the commenters mentioned in concept-related acknowledgments (Cronin, 1995). We omit research assistants, editorial support and non-academic commenters (such as industry professionals or central bankers) from the analysis if they are acknowledged as such. If individuals are not thanked for a specific role, we assume they are acknowledged for concept-related help. Similar to Brown (2005), we remove the journal's managing editors of the current and the previous two years from each paper's list of commenters. This prevents a technical overestimation of their importance in the network.[7] We define a paper as having an acknowledgement if it acknowledges at least one commenter (after removing editors) or at least one seminar or at least one conference. If papers report "multiple" or "several" instances of a category (i.e., "audiences at multiple seminars") we assume that number to be two. For all papers we set the count of a category to 0 if it is not reported. The underlying assumption is that since the paper contains an acknowledgement section, it would have reported an event if presented at one. We then consolidate all 18,949 name variants in our database manually.[8] We obtain 14,787 distinct researchers (author or commenter).

Researcher Characteristics For each researcher and each year we define a number of researcher characteristics. We derive them mostly from records in Elsevier's bibliometric database Scopus using code developed by Rose and Kitchin (2019). Unlike other bibliometric databases, Scopus solves the problem of name disambiguation using unique Author profile identifiers on their side. According to our procedure, detailed in appendix C.1, we match 12,092 or 81.65% of the 14,787 researchers.[9] They account for about 93.6% of 48,141 acknowledged comments. Among the researchers with Scopus match are all 6,552 authors and 9,384 out of 12,079 (77.68%) acknowledged commenters. Unmatched researchers drop out of any sample, but remain in the networks.

We estimate gender based on first name, if available, using the genderize.io database.[10] We obtain gender estimates for 98.63% of all matched researchers in our dataset. That is, 207 researchers remain without estimate. These cases are dropped when we investigate gender questions.

For each matched researcher, we record all research articles and their annual citation count in each year.[11] We call these variables "Publications" and "Citations". Combining both the number of publications and the individual count of citations per paper, we measure prolificness using

[7] The vast majority of papers acknowledge the editor of the respective journal. If we calculate an editor's position within the social network of informal collaboration, we bias towards journals with a higher publishing frequency.

[8] 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.

[9] 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 indexes. Many acknowledged commenters do not satisfy these criteria as they are not academics but industry professionals or research assistants not marked as such.

[10] See https://genderize.io/.

[11] The Scopus document types we use are article, review, conference proceeding and short article. These types are used for journal citation counts.

the Euclidean index of citations called "Euclid".[12] The computation is as follows: For each year t, count all citations to each of researcher i's q papers published until and including t, then take the square root of the sum of the squared citation counts. That is, if $c_{k,t}$ is the total citation count for paper k until t, then the $\text{Euclid}_{i,t}$ of researcher i in year t is:

$$Euclid_{i,t} = \sqrt{\sum_{k=1}^{q_{i,t}} c_{k,t}^2} \tag{1}$$

In a next step, we define experience as the number of years between the first publication and the year of publication of the paper (that either the author published or the commenter is acknowledged on). If the first publication lies in the future (e.g. for commenting PhD students), we set experience to 0. Finally, we record every matched researcher's affiliation using any given year's publications. This allows us to identify colleagues.

Network Measures Using acknowledgements of papers and authorship information, we construct networks of informal collaboration and co-author networks. In the network of informal collaboration, two researchers are connected with a weighted directed link whenever one acknowledges the other 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 since the directionality allows to trace whom researchers acknowledge and by whom. We compare the network of informal collaboration with a co-author network which connects researchers with an undirected weighted link whenever they have co-authored a paper in our dataset.

We construct thirteen networks of informal collaboration and thirteen co-author networks for all $t \in \{1999, 2001, \dots, 2011\}$ according to the same principles.[13] The adjacency matrix G_t represents the network of informal collaboration. Its elements g_{ij} indicate the number of times j acknowledges i in $t, t-1, t-2$, accounted for the number of authors on a paper. That is, for each paper with k authors that acknowledges j, g_{ij} increases by $1/k$. Likewise the adjacency matrix H_t for the co-author network contains elements $h_{kj} = h_{jk}$ indicating the number of joint publications of j and k in $t, t-1, t-2$. In the co-author networks link weight reflects the frequency of joint papers.

Networks can have multiple components and two researchers belong to the same component if there exists an alternating sequence of researchers and links between them. This sequence is called a path. 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 it is also large compared to the rest of the network (Jackson, 2014). Some measures are component-specific, in which case we report the measure only for the largest component. The length of the shortest path between two researchers is their distance. The diameter of a component is the maximum of all shortest paths.

[12] Perry and Reny (2016) show that this index possesses desirable properties that other indices (such as the h-index) do not possess, such as depth relevance, scale invariance, and directional consistency.

[13] We therefore omit the time index when no confusion can arise.

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

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

Clustering refers to the connectedness of a researcher's collaborators: How often do the collaborators of a researcher collaborate with each other? Formally, a researcher i's clustering coefficient is the share of collaborators that are collaborating with each other over the number of possible pairs. The set of collaborators of i is the neighborhood $\mathcal{N}_i(G) = j : g_{ij} > 0$. For the directed networks of informal collaboration, clustering is defined as:

$$clustering(G)_i = \frac{|\{g_{jk} : v_j, v_k \in \mathcal{N}_i, g_{jk} \in G\}|}{\text{degree}(i)(\text{degree}(i)-1)}, \tag{3}$$

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

$$clustering(H)_i = \frac{2|\{h_{jk} : v_j, v_k \in \mathcal{N}_i, h_{jk} \in H\}|}{\text{degree}(i)(\text{degree}(i)-1)}. \tag{4}$$

The average clustering normalizes the sum of all clustering coefficients by the number of network members n:

$$avg.\ clustering = \frac{1}{n} \sum_{j \in G} \text{clustering}_j \tag{5}$$

Network Centralities We compute three network centralities: Out-Degree, eigenvector centrality, and betweenness centrality. For eigenvector centrality and betweenness centrality we use ranks because scores are not comparable across networks and years. For technical reasons, we compute all centralities (except out-degree) in each network's giant component only. Researchers in the other components receive a centrality of 0. This is because the computation of the centralities relies on paths, which means that centralities are not comparable across components. For example, if researcher i belongs to a small network component, all other researchers are fairly close. In contrast, a researcher in a large component might have a potentially much smaller centrality because many other researchers are far away.

Degree and out-degree are two simple and informative measures. It counts (outgoing) links from a researcher. In networks of informal collaboration, it counts how many authors acknowledged someone, and in the co-author networks, which are undirected networks, it is the number of distinct co-authors. More formally (out-)degree is the size of the neighborhood $\mathcal{N}_i(G)$:

$$\textit{(out-)degree}_i = \left|\mathcal{N}_i(G_i)\right|, \tag{6}$$

Researchers that are not acknowledged receive an out-degree of 0. Remember that within the co-author networks $g_{ij} = g_{ji}$.

We are also interested in reach and influence beyond the immediate neighborhood. Since degree is limited to that neighborhood, we also include eigenvector centrality (Bonacich, 1987). Ballester et al. (2006), Hojman and Szeidl (2008) and Elliott and Golub (2019), 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 the provision of public goods such as knowledge, because the effort brought forward in equilibrium corresponds to someone's eigenvector centrality. The eigenvector centrality of researcher i is the weighted sum of collaborators, where weights correspond to their eigenvector centralities, normalized by some constant λ:[14]

$$\textit{eigenvector}_i = \frac{1}{\lambda} \sum_{j \in \mathcal{N}_i} \text{eigenvector}_j \tag{7}$$

Since eigenvector centrality focuses only on connectivity and influence, but remains silent about the importance of a researcher for knowledge flows, we also study betweenness centrality (Freeman, 1978). Betweenness centrality is often used to measure the individual influence on information flows within a network (Jackson, 2014).

A high betweenness central researcher is someone who bridges two otherwise disparate or sparsely connected clusters. They could hold authority over information flow between them, or control collaboration between them. Formally, it is the frequency with which a researcher is on the shortest path $\sigma(j,k)$ between any two researchers j,k:

$$\textit{betweenness}_i = \sum_{j,k \in G} \frac{\sigma(j,k|i)}{\sigma(j,k)} \tag{8}$$

Finally, we add an indicator for membership in the largest component of the respective co-author networks.

Samples There are two main samples we use in this study, the "paper sample" and the "person sample". The paper sample consists of 5,769 observations where each observation is a paper

[14]The intuition behind this definition is that the more important collaborators are, the more important the focal researcher is. By virtue of the Perron-Frobenius theorem, the vector of eigenvector centralities is the eigenvector of the leading eigenvalue of G when λ satisfies the following condition: $G\mathbf{b} = \lambda\mathbf{b}$, i.e., when λ is the leading eigenvalue.

with an acknowledgement (at least one commenter or one seminar or one conference acknowledged).[15] Table 1 presents summary statistics and Table A1 (in the appendix) the corresponding correlation coefficients. Variables in this sample include measures of academic success, paper characteristics (which include characteristics of authors), informal collaboration and centralities of both authors and commenters. Remember that if some forms of informal collaboration are not mentioned (i.e., no conferences acknowledged) but there exists an acknowledgement, we set its count to zero.

Table 1: Summary statistics for the paper sample.

	N	Mean	Median	Std.Dev.	Min	Max
Academic success						
Total citation count	5769	129.6	58	249.63	0	6149
Top publication	5769	0.6	1	0.50	0	1
Paper Characteristics						
# of pages	5769	25.9	25	10.27	1	80
# of authors	5769	2.1	2	0.84	1	6
Auth. total Euclid	5769	186.4	60	396.58	0	8123
Informal collaboration						
# of seminars	5769	3.6	2	4.38	0	32
# of conferences	5769	1.8	1	1.98	0	23
# of commenters	5769	8.3	7	6.56	0	58
Com. total Euclid	5769	1358.4	654	1976.98	0	19459
Authors' centralities						
Auth. giant (co-author)	5769	0.1	0	0.29	0	1
Auth. eigenvector (co-author)	5769	0.0	0	0.05	0	1
Auth. betweenness (co-author)	5769	0.0	0	0.07	0	1
Auth. giant (informal)	5769	0.7	1	0.47	0	1
Auth. eigenvector (informal)	5769	0.0	0	0.04	0	0
Auth. betweenness (informal)	5769	0.0	0	0.01	0	0
Commenters' centralities						
Com. giant (informal)	5769	0.8	1	0.42	0	1
Com. eigenvector (informal)	5769	0.1	0	0.15	0	2
Com. betweenness (informal)	5769	0.0	0	0.04	0	0

Notes: Summary statistics for paper sample, where the unit of observation is a published research paper.

The person sample consists of 57,919 matched researcher-year observations and combines all researcher-related information. These include network centralities but also raw counts of informal collaboration measured in a three-year window.

Table 2 presents summary statistics for the person sample. The average observation has a Euclidean index of 99.2, and has published about 13 publications during 12 years of academic experience, which have garnered 270 citations. Measures of informal collaboration reflect engagement in 3-year windows. The average observation has been acknowledged 2 times by 4 different authors (out-degree). We use ranks instead of centrality scores for eigenvector and betweenness centrality to establish comparability across networks. Table A2 (in the appendix) presents corresponding Spearman correlation coefficients. It is noteworthy how weakly any

[15] Note there are 5,784 papers with acknowledgements in the study period. The missing papers, which are all published in the JMCB, unfortunately are not indexed in Scopus as of publication.

Table 2: Summary statistics for the person sample.

	N	Mean	Median	Std.Dev.	Min	Max
Researcher Characteristics						
Euclid. Index	57,919	99.2	28	257.06	0	9404
Publications	57,919	13.2	8	16.80	0	361
Citations	57,919	270.1	52	806.87	0	34706
Female	57,919	0.2	0	0.36	0	1
Experience	57,919	12.0	10	10.08	0	70
Network Centralities						
No. of Thanks	57,919	2.0	1	3.41	0	104
Out-Degree	57,919	4.0	2	6.64	0	138
Eigenvector centrality rank (informal)	52,825	2418.5	2195	1650.63	1	6933
Betweenness centrality rank (informal)	52,825	2260.5	2133	1485.62	1	5540
Giant membership (co-author)	57,919	0.048	0	0.21	0	1
Degree (co-author)	26,083	1.9	2	1.40	0	21
Eigenvector centrality rank (co-author)	2761	222.1	173	184.05	1	654
Betweenness centrality rank (co-author)	2761	167.6	144	120.14	1	333

Notes: Summary statistics for the person sample, where the unit of observation is the combination of researcher i and year t.

centrality rank correlates with author characteristics, as no coefficient surpasses 0.14 in absolute terms.[16] The number of thanks correlates with productivity measures only weakly, too: Of all the Spearman correlations between number of thanks and any of the author metrics, the highest is with the Euclidean index of citations and equals 0.38.[17]

3 The Nature of Informal Collaboration in Financial Economics

Acknowledging intellectual help by others has become the norm in financial economics. Out of 6,597 papers published in the six journals during 1997 and 2011, 5,784 (≈ 90%) acknowledge seminar presentations, conference participation, or researchers, or combinations of these.[18] More than half of all papers report all three forms of informal collaboration (Figure 1), and less than a quarter report only one form. Figure 2 shows that while the number of papers without acknowledgement section is virtually zero in most journals by 2011, only the Journal of Banking and Finance has a high number of papers without acknowledgements (72 out of 262 published papers in 2011).

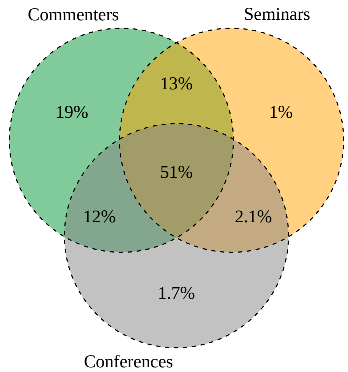
Top journal publications acknowledge informal collaboration more often and report a higher intensity thereof. In fact, the ranking of the six journals by journal impact factor reflects well how they would rank according to the average intensity of informal collaboration. Conditional on having an acknowledgement, the average top journal publication acknowledges almost twice as many commenters as the average non-top journal publication (10.7 vs. 6.2). It is also pre-

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

[17] We find the same pattern when looking at different periods individually.

[18] The remaining papers may, although rarely, acknowledge the editor, anonymous referees, funding, data exchange, and research assistance only.

Figure 1: Share of papers jointly acknowledging researchers by name, seminars and conferences.



Notes: Venn diagram shows the share of papers that acknowledge commenters, seminars and conferences, and combinations thereof. Figure uses the paper sample.

sented more than twice as often at seminars (6.5 vs. 3.2). Interestingly, the number of conferences is almost the same (2.9 vs. 2.2).[19]

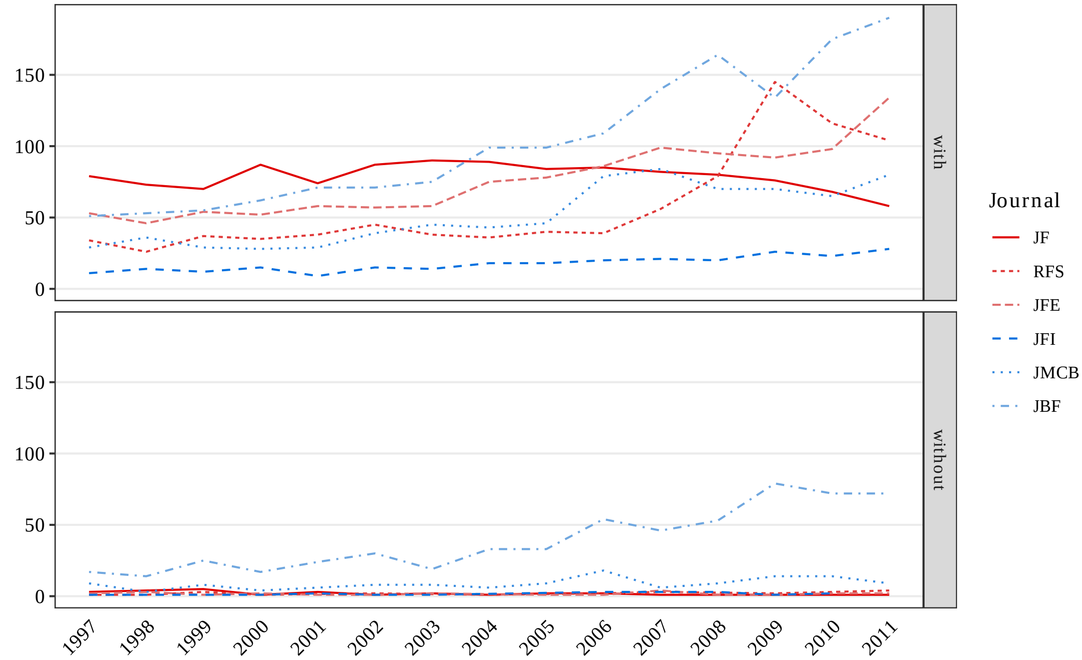
There is more informal collaboration reported over time, but there are also more authors on a paper. Figure 3 thus normalizes collaboration measures by the number of authors on a paper. Papers are not presented in more seminars per author and there are not more commenters per author. Only the number of conferences per authors increases significantly: papers published in 1997–2001 report 0.6 conferences per author on average, while in the 2007–2011 period this number increased to 0.85.

How much authors engage in informal collaboration is partly a question of their academic cohort. Figure 4 compares the mean values of the yearly average of an author's author-normalized number of acknowledged commenters, seminars and conferences, and also the number of co-authors across cohorts. That is, for each paper we count the acknowledged commenters, seminars and conferences, and divide the value by the number of authors, which we then assign to each author. This number is averaged over all authors from the same cohort, which we define as the decade they started publishing in. If an author publishes multiple times per year, we take the average. Authors who start publishing later in our dataset acknowledge significantly more commenters than those who started earlier. The same relationship holds for seminars and conferences. Interestingly, authors from the 2000s have fewer distinct co-authors (in our dataset) than those from the 1990s, who have less than those from the 1980s.

For authors, it seems, informal collaboration is a relevant part of the production of a paper, but it's unclear whether informal collaboration matters at all. Laband and Tollison (2000) show that for 251 featured articles published in the Review of Economics and Statistics during the years 1976–1980, there is a positive correlation between the number of commenters and the subse-

[19] The averages of the non-top journals are driven by the Journal of Financial Intermediation, whose average values are regularly between those of the top journals and the other non-top journals. One plausible explanation is that the JFI publishes a high amount of papers prepared for and rejected by the top journals.

Figure 2: Papers with and without acknowledgements, by journal and year.



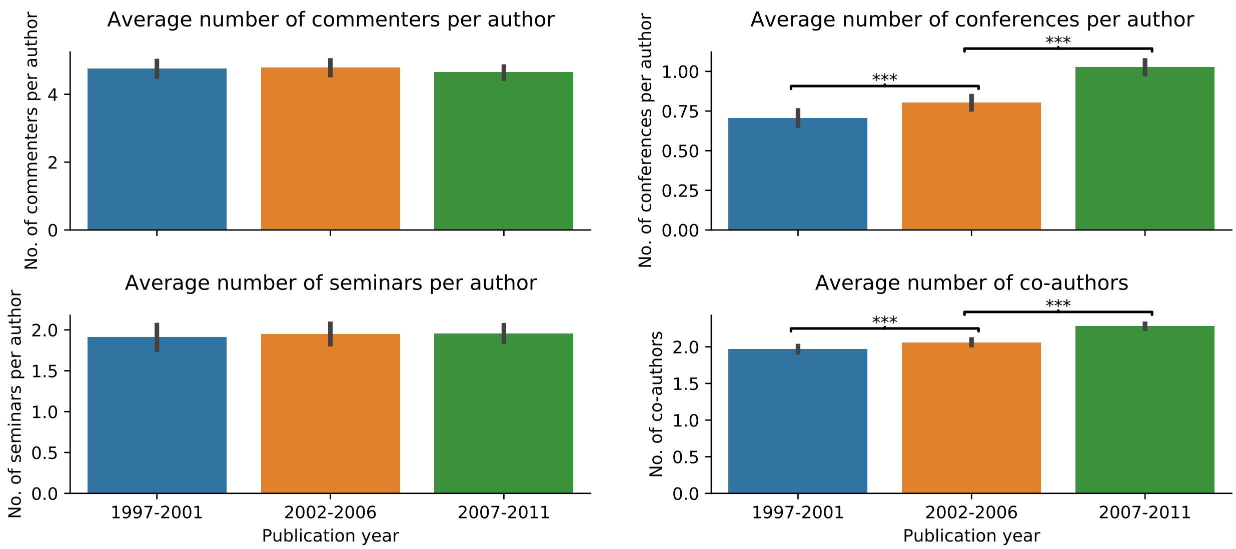
Notes: Graph shows the number of papers with (upper panel) and without acknowledgments section (lower panel) for each year, by journal. An acknowledgement may contain named researchers mentioned for feedback, advice and discussion (unless she's the journal's managing editor), seminars or conferences. See Section 2 for variable and sample definition.

quent citation count. Commenters that are cited more matter even more. While this study leaves endogeneity issues aside, Brown (2005) exploits submissions to three accounting journals. Studying 256 papers, he finds that papers that have been presented more often in seminars have a higher acceptance probability. Both studies (which we replicate in Appendix B.2 using our sample) have a relatively small sample size. Using our paper sample, we want to shed light on the relationship between informal collaboration and academic success. Informal collaboration might be positively associated with a paper's success because feedback can improve the quality of the paper, while presentations can help in the dissemination before and after publication. Thus, we estimate the following regression model:

$$\begin{aligned}\text{Success}_p =& \boldsymbol{\alpha}_1 \text{Paper Characteristics}_{p,t-1} + \beta_1 \text{No. of seminars}_p + \beta_2 \text{No. of conferences}_p \\ &+ \beta_3 \text{No. of commenters}_p + \beta_3 \text{Commenter Quality}_p + \mathbf{D}_{\text{Journal}_p} + \mathbf{D}_t + \varepsilon_p,\end{aligned} \tag{9}$$

We measure "Success" in four ways: the citation count of paper p as of December 2020, the citation count after 5 years, the citation count after 10 years, and by an indicator variable that equals 1 if the paper got published in a top three finance journal. "Paper Characteristics" contains the number of pages, dummies for author group size, and the authors' total Euclidean index of citations in the year before publication. These variables capture paper quality past publication. "Commenter Quality" contains the commentator's total Euclidean index of citations defined in

Figure 3: Collaboration intensity by cohort of papers.

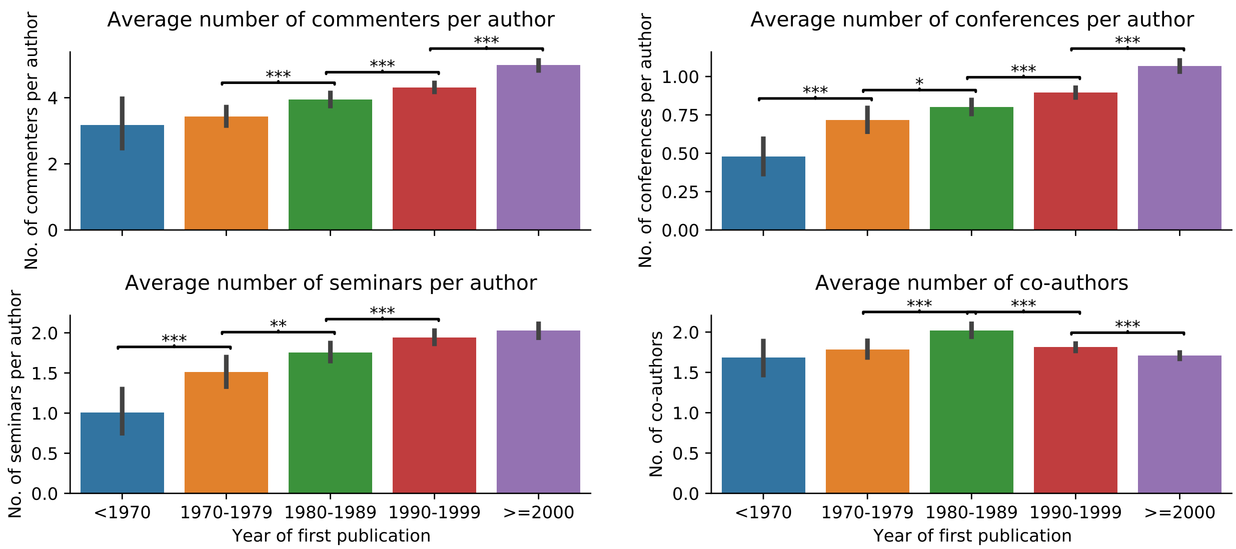


Notes: Figures show the mean number of acknowledged author-normalized amount of informal collaboration, by cohort. Cohort is inferred form the publication year of the paper. Vertical bars indicate 95% confidence intervals. Horizontal parenthesis indicate statistically significant differences, with ***, ** and * indicating statistical significance at the 1, 5 and 10 percent level, respectively.

Equation. We also include journal fixed effects, except for top journal publication status. They pick up journal-specific effects such as popularity, topic, and quality of the editorial process (assuming these are time-independent). To allow for more flexibility, author group size enters as dummies and not as a continuous variable. Without claiming causality, Table 3 shows that all variables are positively associated with paper success. The only exception is the number of conferences, which does not seem to matter for the 10-year and total citation count. The relationship with the number of acknowledged commenters is of particular interest. One of the facts that the co-author networks literature established is that teams are more productive and influential than solo-authors (Medoff, 2003; Wuchty et al., 2007; Ductor, 2015). Yet, it is unclear whether the definition of "teams that make up an influential paper" might extend to informal collaborators as well. We continue this thought towards the end of the paper.

Overall, we do find more informal collaboration over time, but no significant differences across papers, young authors are driving the trend. Papers in top journals are presented at the same number of conferences as papers in non-top journals, but both are presented at more conferences over time. At the same time, authors go to more conferences the younger they are. None of this is to the detriment of other forms of informal collaboration. Both trends could be explained by more informal collaboration in the 2000s as compared to previous decades, for example due to more job presentations or preparation conferences. Better opportunities to attend conferences (either via cheaper transportation, more funds, or more conferences overall)

Figure 4: Collaboration intensity by cohort of author.



Notes: Figure shows the mean number of a person's yearly average number of author-normalized amount of informal collaboration, as acknowledged on published papers, by cohort. Cohort is inferred form the year of first publication of an author. Vertical bars indicate 95% confidence interval. Horizontal parenthesis indicate statistically significant differences, with ***, ** and * indicating statistical significance to the 1, 5 and 10 percent level. Figures use the paper sample. See Section 2 for variable and sample definition.

may be another explanation, if the propensity to travel depends somehow on a person's experience. Exploring these possibilities may help to understand the bigger picture-question whether more informal collaboration improved the quality of papers or science in general.

Table 3: Citation count and informal collaboration.

	Top publication	Total citation count	5-year citation count	10-year citation count
	logistic		*negative binomial*	
	(1)	(2)	(3)	(4)
# of seminars	0.184***	0.009***	0.011***	0.010***
	$p = 0.000$	$p = 0.010$	$p = 0.0003$	$p = 0.004$
# of conferences	−0.037	0.004	0.019***	0.011
	$p = 0.106$	$p = 0.582$	$p = 0.005$	$p = 0.111$
# of commenters	0.037***	0.013***	0.011***	0.013***
	$p = 0.00005$	$p = 0.00001$	$p = 0.00001$	$p = 0.00001$
Com. total Euclid	0.0004***	0.00004***	0.00003***	0.00003***
	$p = 0.000$	$p = 0.0002$	$p = 0.00002$	$p = 0.0002$
Constant	−3.389***	4.944***	2.189***	3.392***
	$p = 0.000$	$p = 0.000$	$p = 0.000$	$p = 0.000$
Paper Characteristics	✓	✓	✓	✓
Author group size fixed effects	✓	✓	✓	✓
Publication year fixed effects	✓	✓	✓	✓
Journal fixed effects		✓	✓	✓
N	5,769	5,769	5,769	5,769
Akaike Inf. Crit.	4,928.259	64,645.970	43,779.250	57,133.120

Notes: Reported coefficients are marginal effects. ***, ** and * indicate statistical significance to the 1, 5 and 10 percent level, respectively. Table uses the paper sample. See Section 2 for sample and variable definition.

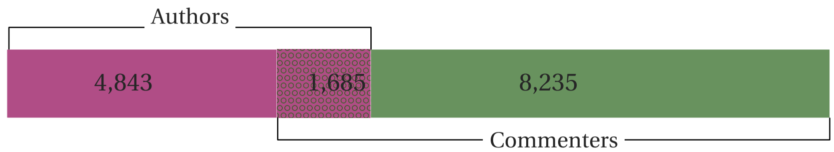
4 Bilateral Informal Collaboration

4.1 Who gets Acknowledged?

The production of the 6,597 papers in our dataset involved many more researchers than just the authors. A total of 8,235 researchers appear solely as commenters but not as authors. On the other hand, most of the 6,552 authors are authors but not commenters of papers in our dataset (Figure 5). This relationship holds for any given year and irrespective of generational differences. Only a minority of 1,685 researchers both comments on and publishes the 6,597 papers in our dataset. On average, we find 1.26 pure commenters for every author in our dataset. Within the three top journals this ratio increases to 1.63.

That a large fraction of commenters does not publish (in our dataset) might have multiple causes. The first is the relatively small number of 6,597 articles we study, compared to about 95,800 publications published by 12,092 researchers during the 1997–2011 period. At the same time, this implies that authors of papers in our dataset acknowledge researchers whose main field is not necessarily finance. On a high level, we find that pure commenters (with matched Scopus ID) are more likely to be mainly active in fields other than economics or business: Economics or business is the main field of 67% pure authors, compared to 41% of pure com-

Figure 5: 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.

menters.[20] In particular, the social sciences field–which includes law, political science, and others–is more relevant in the population of pure commenters. This speaks to the idea that authors source knowledge they don't possess from their social capital (Li et al., 2013). Thus, we think the real surprise is somewhere else: It is not the high number of commenters that are not authoring in our dataset, but the high number of authors that are not commenters in our dataset.

Therefore, we like to explore reasons why some people are acknowledged while others aren't. It is reasonable to assume that authors turn to researchers that are both visible and productive, or that have more experience in publishing and positioning papers. We thus correlate a researcher's number of published papers and the total number of acknowledgements received with that person's characteristics. This ignores possible (and important) bilateral characteristics, such as being colleagues or co-authors. Using the person sample, we estimate the following regression model:

$$\begin{aligned} \text{Collaboration}_{i,t} = & \beta_0 + \beta_1 \text{Euclid}_{it} + \beta_2 \text{Publications}_{it} + \beta_3 \text{Citations}_{it} + \\ & \beta_4 \text{Experience}_{it} + \beta_5 \text{Experience}^2_{it} + \beta_6 \mathbf{D}_{\text{Female}_i} + \epsilon_{it}, \end{aligned} \tag{10}$$

where "Collaboration" is either the "No. of Thanks" or "Out-Degree", i.e. how many papers or how many authors have acknowledged someone. Experience enters the regression explicitly so that we can obtain and report a coefficient. We estimate a negative binomial regression to comply with the count data nature of both variables. Standard errors clustered on the researcher level to account for unobserved heterogeneity. Table 4 reports the results as marginal effects. Their interpretation is that of a percentage increase of the dependent variable if the variable increases by one unit from sample mean, holding all other variables constant at their mean.

All characteristics correlate statistically significantly with the dependent variables. This is independent of whether we include year fixed effects to detect variation across years, or not. Female researchers are acknowledged less often and have a lower out-degree, even at the same level of academic productivity and experience as a comparable male counterpart. On average, female researchers are acknowledged by 19.1% fewer papers (~ 0.4 papers) and by 15.1% fewer authors (~ 0.6 authors) over a three-year period. This finding adds to recent debates regarding the role and perception of female researchers in academia (Chari and Goldsmith-Pinkham, 2017; Sarsons, 2017; Hengel, 2020; Ductor et al., 2018). However, the finding warrants further research to

[20] We define fields via the All Science Journal Classification (ASJC) provided by Elsevier. Economics refers "Economics, Finance and Econometrics", while business refers to "Business, Management and Accounting".

trace causes and consequences. Do we see statistical discrimination or discrimination by taste? Is it a sign of sexism in the community? Is it because authors approach them less often? Or do female researchers shy away from offering comments and insights to authors?

Table 4: Negative binomial regression on engagement in informal collaboration.

	No. of Thanks		Out-Degree	
	negative binomial			
	(1)	(2)	(3)	(4)
Euclid. Index	0.001***	0.001***	0.001***	0.001***
	(0.00004)	(0.00004)	(0.00005)	(0.00005)
Publications	−0.005***	−0.005***	−0.005***	−0.005***
	(0.0004)	(0.0004)	(0.0004)	(0.0004)
Citations	0.0004***	0.0004***	0.0003***	0.0003***
	(0.00001)	(0.00001)	(0.00002)	(0.00002)
Female	−0.191***	−0.187***	−0.151***	−0.160***
	(0.014)	(0.014)	(0.014)	(0.014)
Experience	0.068***	0.068***	0.063***	0.063***
	(0.001)	(0.001)	(0.001)	(0.001)
Experience^2	−0.002***	−0.002***	−0.002***	−0.002***
	(0.00004)	(0.00004)	(0.00004)	(0.00004)
Constant	0.118***	0.106***	0.815***	0.885***
	(0.011)	(0.017)	(0.011)	(0.017)
Year fixed effects		✓		✓
N	57,919	57,919	57,919	57,919
Akaike Inf. Crit.	210,795.500	210,764.700	279,153.700	279,079.300

Notes: Standard errors clustered around individual researchers in parenthesis. ***, ** and * indicate statistical significance to the 1, 5 and 10 percent level, respectively. Table uses the person sample. See Section 2 for variable and sample definition.

4.2 Reciprocity and Inter-Generational Transfer

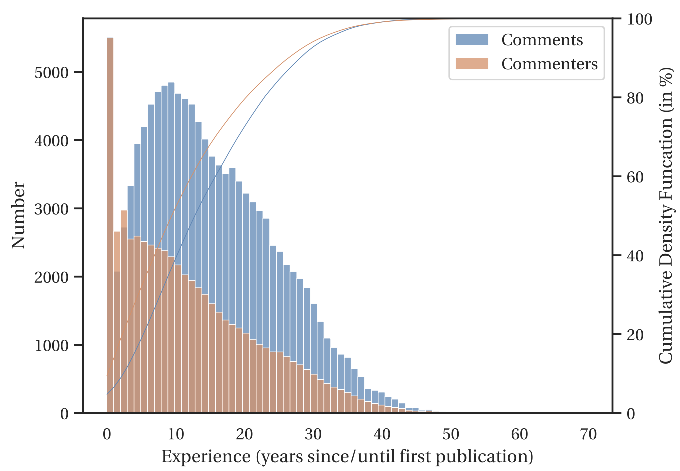
The prevalence of commenting on each others' work raises another question, namely why researchers invest their scarce time to read manuscripts when they do not receive tangible credit for it. One possible explanation is reciprocity: Researchers help other researchers who have helped them in the past or who are likely going to help them in the future (Malmendier et al., 2014). Another explanation is inter-generational knowledge transfer. We investigate both possibilities below. Investigating a third plausible explanation, intrinsic motivation, is not possible with our data alone.

We define three forms of reciprocal relationships: (i) commenting on the work of coauthors, (ii) commenting on the work of commenters, (iii) commenting on the work of colleagues. We treat author and commenter of a paper as colleagues if in the year before publication they report the same affiliation (not counting platform affiliations like the NBER). We rely on Scopus for the definition of affiliations.

Accordingly, we define a paper p with authors κ_p as exhibiting reciprocity if it satisfies one of three conditions: 1) An acknowledged commenter is a co-author of at least one of κ_p; 2) An acknowledged commenter publishes a paper in our dataset and at least one of κ_p is acknowledged, while none of κ_p co-author on that paper; 3) An acknowledged commenter is a colleague of at least one of κ_p in the year before publication. Naturally, for condition 1) or 2) to hold, at least one of the acknowledged commenters must be an author in the dataset.

Our results confirm that reciprocity is prevalent within the bounds of our dataset. We find 5,167 papers where at least one of the author's coauthors is an author in our dataset. There are 5,147 papers where we could observe reciprocity among commenters. Of these 3,179 papers, at least one acknowledged commenter in turn acknowledges at least one of the authors on her own papers. Finally, there are 3,320 papers acknowledging at least one colleague, out of 5,467 papers where we could observe acknowledgments to colleagues. In total, 4,261 papers (about 77.9% of 5,467 papers) fulfill one of the above conditions.

Figure 6: Histogram and CDF for number of commenters and comments, by academic experience.



Notes: Histogram showing the number of commenters (red) 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.

Our estimates of reciprocity should be seen as lower bound. It might be higher with a larger dataset. We only observe commenter links within a set of six journals. As shown above, financial economics is not necessarily the natural domain of all commenters.[21] Some may well acknowledge authors that acknowledged them on papers outside our dataset. Similarly, affiliation

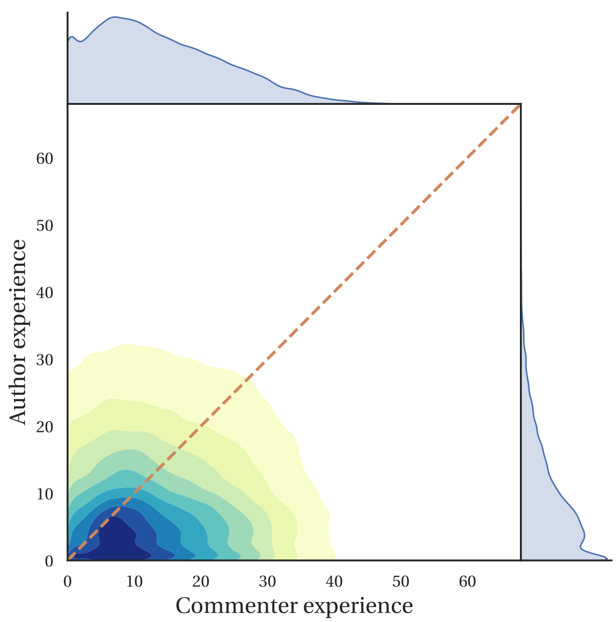
[21] For example, the 2014 Economics Nobel laureate, Lars Peter Hansen, has been acknowledged by more than 20 papers in our dataset, while he didn't author one paper in our dataset. The same holds true for the 1993 Nobel laureate Douglass North, who appears in two acknowledgement sections.

information is sometimes imperfect, posing challenges to our definition of colleagues.[22]

Another explanation for the high engagement in informal collaboration is an implicit generational exchange or transfer of knowledge and experience. Instead of investing all their time in the production of new papers, senior researchers may choose to help young and rising authors. While the range of our data is too short to observe an entire academic generation from graduation to retirement, we can at least confirm two necessary conditions. The first is that commenting is not evenly distributed across a researcher's life-cycle. The other necessary condition is that young authors acknowledge senior authors more often than vice versa. The red histogram and the corresponding CDF in Figure 6 show that the majority of commenters have 3 to 20 years of academic experience when the paper they are acknowledged on is published. The mode positive experience is 2 years. When weighing commenters with the number of given comments the mode positive experience is 9 years (blue histogram and grey CDF).

Consequently, informal collaboration where the author is less senior then the commenter is the norm. Figure 7 exemplifies this. Most of the observations locate right of the 45° line, which would show equal experience.

Figure 7: The distribution of authors' and commenters' experience.



Notes: Figure shows the joint distribution of academic experience of authors (left axis, right marginal plot) and of academic experience of commenters (bottom axis, top marginal plot) via kernel density estimation. Dark areas indicate higher density. Experience is the number of years between first publication and the publication year of the paper that acknowledges the commenter.

[22]Like commenters' names, which are often misspelled, affiliations are often misspelled too. In addition, authors often use abbreviations for universities, in particular when acknowledging feedback in seminars. This makes a manual consolidation procedure necessary.

5 The Network of Informal Collaboration in Financial Economics

5.1 Network Topology

Collaboration of researchers constitutes a network which we use to study information flow and spill-overs at an individual and a global level. To better understand the magnitude and depth of the networks of informal collaboration we contrast them with co-author networks built from the same publications for the same periods.

Table 5 gives an account of the networks' topologies. The 1999 network of informal collaboration is generated from 863 papers published in either 1997, 1998, or 1999 and consists of 3,154 researchers. In comparison, the 2011 network of informal collaboration connects 7,187 researchers that have collaborated on 1,961 papers. The number of researchers in the network grows at about the same rate as the number of annual publications. This is not a trivial result: Rather than acknowledging the existing set of researchers, new authors acknowledge different researchers. Yet, this leaves the question unanswered why the growth rates are so close to each other.

Compared to the co-author networks, the networks of informal collaboration are more inclusive and more connected but less dense. Several figures exemplify this: (i) There are up to 948 distinct components in the co-author networks but no more than 56 components in the networks of informal collaboration; (ii) In the co-author networks, less than a fifth of all researchers are connected within one component, while in the networks of informal collaboration, the largest component includes at least 95% of all researchers. Furthermore, comparing the 1997–1999 and 2009–2011 networks (Figure A1 in the Appendix), we find higher interconnectedness among collaborators involved in publications in top journals; (iii) The average path length and diameter are usually lower in the largest component of the network of informal collaboration than in the co-author network's largest component; and (iv) despite becoming more inclusive, both networks become less dense over time: The density of the network of informal collaboration decreased from 0.0023 in 1999 to 0.0014 in 2011. The growth rate of collaboration, it seems, does not keep up with the increased number of participating researchers.

Only the networks of informal collaboration exhibit small-world properties, but the co-author networks do not. Small-world networks have unique information transfer capabilities (Watts and Strogatz, 1998; Watts, 1999). A small-world network exhibits high clustering, a small average distance, a high number of nodes compared to the number of links, and a giant component. The co-author networks feature high average distance and lack a giant component. The networks of informal collaboration, on the other hand, fulfill all conditions of small-world networks. Even the clustering coefficient, which may appear tiny,[23] is large when compared a random network. If link formation was random, the expected clustering coefficient is approximated by the average number of collaborators divided by the total number of nodes in the network (Goyal et al., 2006). For the 2006 network of informal collaboration, for example, the network with the lowest

[23] It may be that the low clustering is due to the small number of journals. However, constructing networks from the set of the three top journals only results in higher, not lower, clustering.

Table 5: Global network measures.

(a) Networks of informal collaboration

	Overall				Giant				
	Size	Links	Avg. clustering	Components	Size	Density	Avg. path length	Diameter	rho
1999	3154	10,705	0.1	31	3010	0.0023	4.59	12	0.50***
2000	3314	11,327	0.103	35	3138	0.0022	4.63	13	0.48***
2001	3478	11,855	0.12	36	3308	0.0021	4.64	13	0.49***
2002	3649	12,629	0.108	36	3482	0.002	4.74	14	0.49***
2003	3887	13,914	0.106	32	3736	0.0019	4.71	13	0.50***
2004	4220	15,352	0.104	34	4019	0.0018	4.64	13	0.52***
2005	4533	16,959	0.104	31	4400	0.0017	4.68	13	0.49***
2006	4874	18,180	0.088	36	4732	0.0016	4.75	14	0.49***
2007	5381	21,390	0.092	37	5237	0.0015	4.76	14	0.51***
2008	5879	24,216	0.101	40	5698	0.0015	4.70	14	0.53***
2009	6409	28,815	0.105	45	6191	0.0015	4.59	14	0.53***
2010	6827	30,960	0.097	53	6580	0.0014	4.63	14	0.51***
2011	7186	33,630	0.103	56	6932	0.0014	4.55	14	0.52***

(b) Co-Author networks

	Overall				Giant				
	Size	Links	Avg. clustering	Components	Size	Density	Avg. path length	Diameter	rho
1999	1210	972	0.375	486	33	0.0909	3.64	8	0.37**
2000	1267	1028	0.371	500	53	0.0581	5.16	13	0.30**
2001	1366	1129	0.37	531	56	0.0494	6.31	15	0.36***
2002	1460	1210	0.362	561	61	0.0443	6.26	15	0.35***
2003	1511	1281	0.369	567	78	0.0363	6.12	15	0.24**
2004	1668	1439	0.384	607	68	0.0435	6.14	14	0.35***
2005	1798	1585	0.408	648	134	0.0217	8.36	20	0.28***
2006	2059	1808	0.412	747	85	0.0356	6.91	17	0.34***
2007	2312	2210	0.453	758	274	0.0119	10.06	26	0.17***
2008	2620	2610	0.476	814	138	0.0227	6.81	15	0.39***
2009	2873	2986	0.489	836	593	0.0054	12.26	34	0.15***
2010	3075	3150	0.491	910	567	0.0056	14.59	48	0.14***
2011	3213	3306	0.484	948	654	0.0049	11.05	28	0.22***

Notes: Table presents global network statistics for all three-year networks of informal collaboration and of co-author network. Each network of informal collaboration connects researchers that have that collaborated formally (co-authoring) or informally on papers published in year t, $t-1$ or $t-2$. Each co-author 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)) 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 (Equation (5)). ρ is the Spearman rank correlation coefficient between all researchers' betweenness centrality (Equation (8)) and Eigenvector centrality (equation (7)) in the largest component, with ***, ** and * indicating statistical significance to the 1, 5 and 10 percent level.

average clustering, the expected clustering coefficient is 0.0016, roughly 50 times lower than the actual value (0.88). Thus all networks of informal collaboration exhibit small-world properties.

Finally, we would like to point out that betweenness centrality and eigenvector centrality do not strongly correlate. This is indicated by the Spearman correlation coefficient ρ between eigenvector centrality and betweenness centrality, which never exceeds 0.53. This tells us that 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).

5.2 Centrality and Academic Productivity

Being connected to central researchers matters to authors since it enables access to knowledge and ideas traversing the network. This section shows that our data on informal collaboration contains valuable information above what has been used in the literature. Notably, Ductor et al. (2014) show that an economist's future productivity can be forecasted using variables derived from a co-author network. More specifically, the productivity of current coauthors of a researcher as well as the researcher's network centrality measures contain information about her future productivity. One plausible underlying mechanism is that researchers become more productive when they have access to information traversing the network. Such forecasts are relevant for first-time hiring decisions, which for job-market candidates are often based on a small number of signals, the most important of which is the job market paper. Ductor et al. (2014) thus state that hiring departments can improve their information about a job applicant by looking at network centralities of her coauthors.

We repeat this exercise with our data. The aim is to predict the log output of the next three years, where output equals the journal quality weighted number of publication. Predictors include recent past output (log of output of the current and the previous four years), and centralities as well as collaborator measures derived from both the co-author network and the network of informal collaboration. To compare the prediction to the benchmark we use the relative root-mean-square error (RMSE) decrease. Appendix B.1 discusses the empirical setup in greater detail and defines all variable more formally.

Table 6 presents the result of this exercise. Using recent past output as predictor the RMSE decreases by 15.00%. The improvement is in line with the original study which reports 15.72%. However, recent past output might not always be observable, for example for job market candidates. But variables derived from the co-author network are available. The regression using them increases prediction accuracy over the benchmark by 10.00%. Variables from the novel network of informal collaboration alone increase prediction more, namely by 12.50% over the benchmark. Combining variables derived from both networks increases prediction accuracy by exactly 15.00% over the benchmark, which is the same as using recent past output alone. The last regression combines all variables.

Taken together, the exercise shows the networks contain a small amount of information on researchers with publications, and a larger amount for researchers without publications. Specifi-

Table 6: Comparison of forecasts of researcher productivity akin to Ductor et al. (2014).

	Adj. R^2	RMSE	RMSE Differential
Benchmark	0.12	1.20	
Recent past output	0.36	1.02	15.00***
Author network variables	0.29	1.08	10.00***
Commenter network variables	0.33	1.05	12.50***
Auth. net. and com. net. variables	0.37	1.02	15.00***
All	0.41	0.98	18.33***

Notes: 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 model "Benchmark". 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. ***, ** and * indicate statistical significance to the 1, 5 and 10 percent level in Diebold-Mariano tests, respectively. Table uses all author observations from the person sample. Variable construction corresponds to Ductor et al. (2014).

cally, networks of informal collaboration contain information not embedded in corresponding co-author networks, because they are not merely a proxy for informal collaboration.

5.3 Central Commenters and Paper Success

Not only academic careers but also academic papers could benefit from collaboration with central researchers. Authors gain access to knowledge and ideas traversing the network. In this section, the unit of analysis shifts to the individual paper. Section 3 establishes a positive association between informal collaboration and paper success. Starting from this observation, we now ask whether the centrality of authors and acknowledged commenters in the network of informal collaboration contains information about the paper's scientific impact. Specifically, we want to know whether it explains more than measures of the author's and commenters' past academic productivity. We expect this to be the case since connectedness in a network provides access to information in it, as well as to skills, insights, and capacities the co-authors themselves might not possess.

Similar to model (9) above, we model the success of paper p published in year t as:

$$\begin{aligned}\text{Success}_p =& \boldsymbol{\alpha}_1 \text{Paper Characteristics}_{p,t-1} + \boldsymbol{\alpha}_2 \text{Informal Collaboration}_{p,t-1} \\ &+ \boldsymbol{\beta}_1 \text{Author centrality}_{p,t-1} + \boldsymbol{\beta}_1 \text{Commenter centrality}_{p,t-1} + \mathbf{D}_{\text{Journal}_p} + \mathbf{D}_t + \varepsilon_p,\end{aligned} \tag{11}$$

where we measure "Success" as before in four ways: by the citation count of paper p in December 2020, by the citation count after 5 years, by the citation count after 10 years, and by an indicator variable that equals 1 if the paper was published in a top three finance journal. "Paper Characteristics" contains the number of pages, dummies for author group size, and the authors' total Euclidean index of citations in the year before publication. "Informal Collaboration" contains all counts of acknowledged informal collaboration and the commenters' total Euclidean index of citations in the year before publication. "Author centrality" is an indicator for membership in the network's largest component, the authors' sum of betweenness centralities (equation (8)) and the sum of eigenvector centralities (equation (7)). These three centralities are computed both in the network of informal collaboration and for comparison in the co-author network. We use the networks corresponding to $t-1$ to prevent that links observed on p influence the network position of either authors or commenters. Unless we look at citations after a specific number of years, we introduce publication year fixed effects to account for a varying number of years to gather citations.

Table 7 reports estimation results for model (11) using top journal status as dependent variable, and Table 8 using citation counts. Both tables report marginal effects.

Table 7: Top publication status and centrality in the network of informal collaboration.

	Top publication				
	logistic				
	(1)	(2)	(3)	(4)	(5)
Auth. giant (co-author)		1.266^{***}			0.403^{**}
		$p = 0.000$			$p = 0.027$
Auth. eigenvector (co-author)		-2.673^{***}			-2.920^{***}
		$p = 0.002$			$p = 0.002$
Auth. betweenness (co-author)		0.245			0.504
		$p = 0.761$			$p = 0.607$
Auth. giant (informal)			0.984^{***}		0.674^{***}
			$p = 0.000$		$p = 0.000$
Auth. eigenvector (informal)			33.966^{***}		20.199^{***}
			$p = 0.000$		$p = 0.000$
Auth. betweenness (informal)			-36.854^{***}		-37.325^{***}
			$p = 0.000$		$p = 0.000$
Com. giant (informal)				0.849^{***}	0.752^{***}
				$p = 0.000$	$p = 0.00000$
Com. eigenvector (informal)				14.742^{***}	12.795^{***}
				$p = 0.000$	$p = 0.000$
Com. betweenness (informal)				-7.850^{***}	-7.345^{***}
				$p = 0.0002$	$p = 0.0004$
Constant	-3.294^{***}	-3.228^{***}	-2.705^{***}	-2.739^{***}	-2.535^{***}
	$p = 0.000$	$p = 0.000$	$p = 0.000$	$p = 0.000$	$p = 0.000$
Paper Characteristics	✓	✓	✓	✓	✓
Informal Collaboration	✓	✓	✓	✓	✓
Author group size fixed effects	✓	✓	✓	✓	✓
Publication year fixed effects	✓	✓	✓	✓	✓
N	5,769	5,769	5,769	5,769	5,769
Akaike Inf. Crit.	5,722.131	5,645.354	5,192.805	4,758.770	4,587.831

Notes: Reported coefficients are marginal effects. ***, ** and * indicate statistical significance to the 1, 5 and 10 percent level, respectively. Table uses the paper sample. See Section 2 for sample and variable definition.

Table 8: Citation count and centrality in the network of informal collaboration.

	Total citation count					5-year citation count	10-year citation count
	Negative binomial						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Auth. giant (co-author)		0.032			−0.031	0.089**	0.069
		$p = 0.562$			$p = 0.573$	$p = 0.050$	$p = 0.159$
Auth. eigenvector (co-author)		0.441			0.530*	0.202	0.302
		$p = 0.123$			$p = 0.062$	$p = 0.401$	$p = 0.245$
Auth. betweenness (co-author)		−0.146			−0.514**	−0.515**	−0.553**
		$p = 0.528$			$p = 0.031$	$p = 0.011$	$p = 0.012$
Auth. giant (informal)			0.133***		0.110***	0.150***	0.141***
			$p = 0.0005$		$p = 0.004$	$p = 0.00001$	$p = 0.00002$
Auth. eigenvector (informal)			0.801**		0.261	−0.053	−0.012
			$p = 0.034$		$p = 0.515$	$p = 0.876$	$p = 0.975$
Auth. betweenness (informal)			7.391***		7.253***	5.103***	5.480***
			$p = 0.00001$		$p = 0.00001$	$p = 0.0002$	$p = 0.0002$
Com. giant (informal)				0.008	−0.019	0.076**	0.027
				$p = 0.860$	$p = 0.685$	$p = 0.022$	$p = 0.440$
Com. eigenvector (informal)				0.258*	0.206	0.651***	0.491***
				$p = 0.054$	$p = 0.135$	$p = 0.00000$	$p = 0.0001$
Com. betweenness (informal)				3.195***	2.847***	0.715	1.619***
				$p = 0.000$	$p = 0.00000$	$p = 0.122$	$p = 0.002$
Constant	5.042***	5.044***	5.102***	5.059***	5.100***	2.472***	3.631***
	$p = 0.000$	$p = 0.000$	$p = 0.000$	$p = 0.000$	$p = 0.000$	$p = 0.000$	$p = 0.000$
Paper Characteristics	✓	✓	✓	✓	✓	✓	✓
Author group size fixed effects	✓	✓	✓	✓	✓	✓	✓
Publication year fixed effects	✓	✓	✓	✓	✓		
N	5,769	5,769	5,769	5,769	5,769	5,769	5,769
Akaike Inf. Crit.	64,748.090	64,750.810	64,688.100	64,661.870	64,629.760	44,001.700	57,209.410

Notes: Reported coefficients are marginal effects. ***, ** and * indicate statistical significance to the 1, 5 and 10 percent level, respectively. Table uses the paper sample. See Section 2 for sample and variable definition.

Without claiming causality, there are three observations we would like to highlight. First, the 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 at least some coefficients for commenters' centralities are statistically significant in all models, and that goodness-of-fit is better in column (5) of each table than in the previous three columns. Thus, looking at who collaborated informally on a paper provides additional information about it.

Second, and most importantly, models with centralities computed in the network of informal collaboration outperform those models with centralities computed in the co-author network. This can be seen from comparison of the goodness-of-fit of columns (2) and (4) of each table, relative to column (1). Additionally, author centralities in the network of informal collaboration are almost always statistically significant when studying citation counts, but this is not always the case for 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 her ability to receive traversing information–better than co-author networks.

Finally, the centrality of authors and commenters matters for publication in crucial ways. Specifically, betweenness centrality and eigenvector centrality of authors and commenters in both networks are correlated in different ways with the academic success of a paper. Authors and commenters with high betweenness centrality in the network of informal collaboration are associated positively with citation count (columns (5) through (7) of Table 8), but negatively with top journal publication (column (5) of Table 7). The only exception is the short-term citation count (column (6)), indicating that betweenness central commenters pay off in the long run. Overall our finding suggests that authors and commenters that connect different communities (high betweenness centrality) are less likely to publish in one of the top journals although their papers are cited more frequently than the average of the journal. This finding is related to Li et al. (2013) who find that in a small network of information systems scholars betweenness central authors are cited more often. On the other hand, 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 receive above-average citations.

5.4 Who is central?

Having shown that collaborating with central colleagues is more beneficial than collaborating with less central colleagues, we finally turn to the question what determines centrality. We strive to understand how centralities in the networks are related to one another and, more importantly, with observable characteristics of researchers. Considering researcher i in year t, we estimate the following empirical model:

$$\begin{aligned}\text{Centrality}_{i,t} =& \beta_0 + \beta_1 \text{Euclid}_{it} + \beta_2 \text{Publications}_{it} + \beta_3 \text{Citations}_{it} + \\ & \beta_4 \text{female}_i + \mathbf{D}_{\text{Experience}_{it}} + \mathbf{D}_t + \epsilon_{it},\end{aligned} \tag{12}$$

where "Centrality" is one of eigenvector centrality rank and betweenness centrality rank. We cluster standard errors on the researcher level to capture unobserved heterogeneity of which there are two sources. One source of heterogeneity is a different frequency in the data: Some researchers appear in all networks, while others only in one network. The other source of unobserved heterogeneity are different individual networks that authors can draw from. Since the dependent variable is a rank, a negative β indicates a positive relationship.

Table 9: OLS regression on eigenvector and betweenness centrality ranks.

	Eigenvector centrality rank		Betweenness centrality rank	
	(1)	(2)	(3)	(4)
Euclid. Index	-0.207	-0.232^{**}	-0.503^{**}	-0.539^{***}
	(0.171)	(0.097)	(0.216)	(0.125)
Publications	-3.634^{***}	-4.822^{***}	-8.605^{***}	-10.377^{***}
	(1.003)	(0.829)	(1.229)	(0.899)
Citations	-0.072	0.074^{**}	-0.005	0.212^{***}
	(0.064)	(0.032)	(0.079)	(0.039)
Female	13.310	-19.787	84.529^{***}	35.174
	(31.133)	(29.191)	(30.165)	(26.815)
No. of Thanks		-101.069^{***}		-150.717^{***}
		(7.124)		(12.052)
Year fixed effects	✓	✓	✓	✓
Experience fixed effects	✓	✓	✓	✓
N	52,825	52,825	52,825	52,825
Adjusted R^2	0.179	0.220	0.210	0.322

Notes: Standard errors clustered around individual researchers in parenthesis. ***, ** and * indicate statistical significance to the 1, 5 and 10 percent level, respectively. See Section 2 for sample and variable definition.

Table 9 presents results of OLS regressions for model (12) with different dependent variables. Researchers are more eigenvector and more betweenness central the more they publish. This is a trivial result, and might stem from our network construction, because publishing more papers increases the chance of occurring multiple times in our dataset.[24] Controlling for the number of times someone is acknowledged, more prolific authors are more eigenvector central while a higher citation count is negatively affecting eigenvector centrality ranks (column (2)). The reason might be that frequently cited authors are simply unavailable except to a small core set of researchers. More prolific researchers are more betweenness central on average (columns (3) and (4)) as well. Female researchers are on average less betweenness central since fewer papers acknowledge them than comparable male counterparts. But their resulting betweenness centrality is on par with that of males with same characteristics that are acknowledged as often (columns (3) and (4)).

For comparison, we estimate a similar regression for measures computed in the co-author networks. Table 10 reports the corresponding results. Researchers with higher citation counts have fewer distinct co-authors, holding constant the number of papers they have published. We find

[24] This influences the network in two ways: First through co-authors and second through acknowledged researchers. As a result, more productive researchers have more links.

Table 10: Regression results for OLS estimation explaining centrality in the co-author networks.

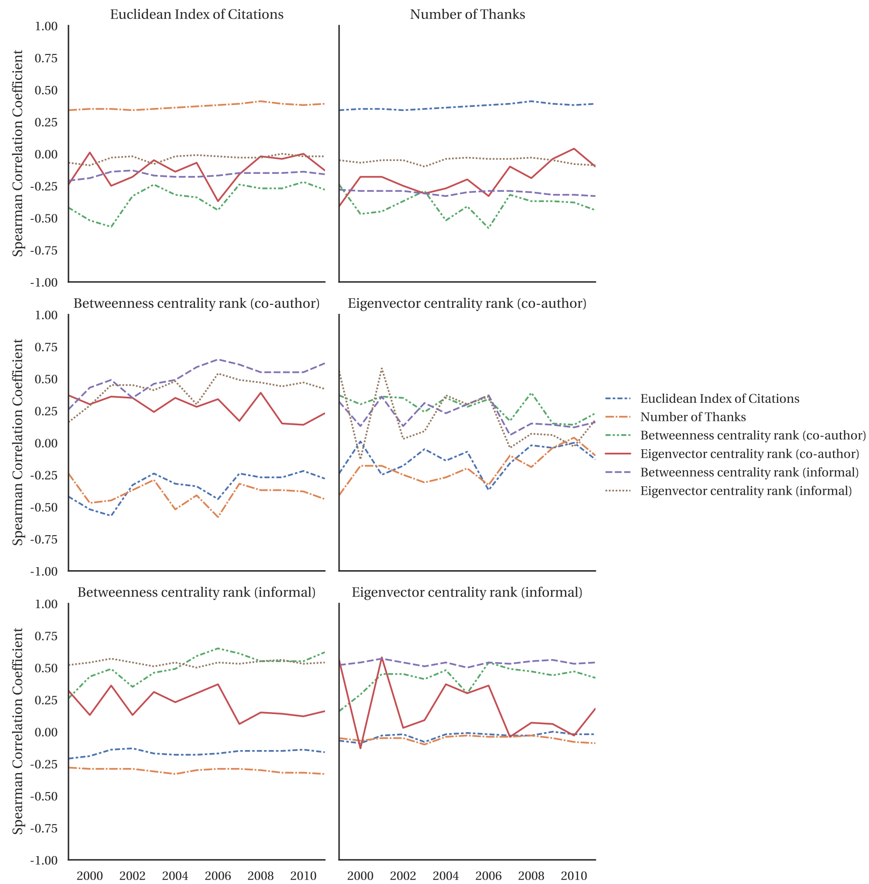
	Degree	Eigenvector centrality rank	Betweenness centrality rank
	(1)	(2)	(3)
Euclid. Index	−0.0002	0.028	0.013
	(0.001)	(0.045)	(0.031)
Publications	0.011***	−0.583	−0.637
	(0.003)	(0.410)	(0.559)
Citations	0.0002	−0.012	−0.015
	(0.0002)	(0.013)	(0.011)
Female	−0.008	−2.413	2.413
	(0.042)	(9.481)	(5.235)
Year fixed effects	✓	✓	✓
Experience fixed effects	✓	✓	✓
N	26,083	2,761	2,761
Adjusted R^2	0.069	0.390	0.552

Notes: Standard errors clustered around individual researchers in parenthesis. ***, ** and * indicate statistical significance to the 1, 5 and 10 percent level, respectively. See Section 2 for variable definition.

no statistically significant relationship between author characteristics and centralities in the co-author network.

Figure 8 shows Spearman correlations over time between six variables to corroborate an important finding: Namely that the most often thanked researchers are not necessarily the most relevant for the flow of information (i.e., high in betweenness centrality), nor the most connected authors (i.e., high in eigenvector centrality). The upper left panel for example confirms that being thanked often and being a very prolific academic were always distinct features, as the Spearman rank correlation coefficient never exceeds 0.4.

Figure 8: Spearman rank correlation coefficients between various variables over time.



Notes: Figures depict Spearman rank correlation coefficients over time for various variables. Figure uses the person sample. See Section 2 for sample and variable definition.

Since the correlations with existing measures of productivity are so low, we argue that centrality in the network of informal collaboration constitutes a new ranking. To make the analysis more palpable, Table 11 hence ranks researchers from the person sample based on their rankings according to the different measures *averaged* over different years. We make the full list for each network available at michael-e-rose.github.io/CoFE. This follows the example of Laband and Tollison (2003) who compile a list of the most often thanked authors from a sample of three general interest economics journals over forty years. Yet, our ranking uses a more homogeneous sample and includes a network view.

Some of the greatest financial economists of our time are prominently featured in the ranking. René M. Stulz, has been acknowledged most often, followed by Jeremy C. Stein and Jay R. Ritter. An interesting observation is that many of the highly ranked researchers are editors. This is despite their removal from the acknowledgment section, if the publication date of a paper falls within the editor's tenure. Apparently authors acknowledge editors even when the paper is not being published in their journal. This highlights the exposed role for editors in the profession (Brogaard et al., 2014). It might also be that authors acknowledge editors even when they rejected the paper.

However, not only editors are acknowledged often or central in the network of informal collaboration: Thus, looking at acknowledgements poses indeed an abundant yet hitherto overlooked source of information of scholarly influence. Discussing the relationship of citations, which matter for academic careers, and acknowledgement, which do not, Cronin (1995, p. 24) notes that “a mere acknowledgement, no matter how influential the acknowledged intellectual contribution is perceived to be, is not treated as equivalent to even the lowest form of citation. A citation of whatever kind is presumed to be more significant (in terms of measurable intellectual debt) than even the most fulsomely worded acknowledgement received, from a mentor, colleague, or peer.” Cronin (1995, p. 22f.) also notes that the absence of informal collaboration from evaluation programs “may be the lack of an acknowledgement database”.

Table 11: Top 30 researchers according to average rankings according to different centrality measures in all co-author and commenter networks.

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

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. See section 2 for variable definition.

6 Outlook

Our financial economics acknowledgements dataset facilitates studying informal collaboration, its effects, consequences and causes. We show that the data contain information not captured by co-authorship networks. The data is novel, high-quality and freely available.

From our work, many avenues for further research open up: One avenue relates to the topology of the network and its impact on and the relation to the profession. The topology of these networks affects the speed of learning and the diffusion of information (Alatas et al., 2016) and is thus of interest for aggregate productivity. There are warning signs, too. Scientific progress depends on the open articulation of thought. Too strong authority (or leadership) may have a negative effect, as it prevents theories from being challenged (Azoulay et al., 2019). Too close networks might result in biased academic reviews (Carrell et al., 2020).

Another avenue is to add 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, 2020). Not all of these need to be performed by authors: Authors can 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 different from the actual authors (Katz and Martin, 1997; Ponomariov and Boardman, 2016). Finally we can gain new insights from understanding the network formation process better: Which role do geography, field, experience and affiliation play? Such insights could guide research policy and also help large, highly innovative firms where groups compete internally and collaborate spontaneously to innovate.

References

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.

Andrews, M. (2019), Bar Talk: Informal Social Interactions, Alcohol Prohibition, and Invention, mimeo.

Azoulay, P., Fons-Rosen, C. and Zivin, J. S. G. (2019), 'Does Science Advance One Funeral at a Time?', *The American Economic Review* **109**(8), 2889–2920.

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

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

Beaver, D. d. and Rosen, R. (1978), 'Studies in Scientific Collaboration Part I. The Professional Origins of Scientific Co-Authorship', *Scientometrics* **1**(1), 65–84.

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

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.

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', *The Accounting Review* **80**(1), 55–83.

Buenstorf, G. and Schacht, A. (2013), 'We need to talk – or do we? Geographic distance and the commercialization of technologies from public research', *Research Policy* **42**(2), 465–480.

Campos, R., Leon, F. and McQuillin, B. (2018), 'Lost in the Storm: The Academic Collaborations That Went Missing in Hurricane ISSAC', *The Economic Journal* **128**(610), 995–1018.

Carrell, S., Figlio, D. and Lusher, L. (2020), Clubs and Networks in Economic Reviewing, mimeo.

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

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

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

de Leon, F. L. L. and McQuillin, B. (2020), 'The Role of Conferences on the Pathway to Academic Impact: Evidence from a Natural Experiment', *Journal of Human Resources* pp. 1116–8387.

Desrochers, N., Paul-Hus, A. and Pecoskie, J. (2017), 'Five Decades of Gratitude: A Meta-Synthesis of Acknowledgments Research', *Journal of the Association for Information Science and Technology* **68**(12), 2821–2833.

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* (2018/07).

Elliott, M. and Golub, B. (2019), 'A Network Approach to Public Goods', *Journal of Political Economy* **127**(2), 730–776.

Freeman, L. C. (1978), 'Centrality in Social Networks Conceptual Clarification', *Social Networks* **1**(3), 215–239.

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.

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., Jiang, L., Thursby, J. and Thursby, M. (2014), 'Specific and general information sharing among competing academic researchers', *Research Policy* **43**(3), 465–475.

Haeussler, C. and Sauermann, H. (2020), 'Division of labor in collaborative knowledge production: The role of team size and interdisciplinarity', *Research Policy* **49**(6), 103987.

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

Hasan, S. and Koning, R. (2019), 'Conversations and idea generation: Evidence from a field experiment', *Research Policy* **48**(9), 103811.

Hengel, E. (2020), Publishing while Female. Are Women held to Higher Standards? Evidence from Peer Review, mimeo.

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

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

Katz, J. and Martin, B. R. (1997), 'What is research collaboration?', *Research Policy* **26**(1), 1–18.

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.

Malmendier, U., te Velde, V. L. and Weber, R. A. (2014), 'Rethinking Reciprocity', *Annual Review of Economics* **6**(1), 849–874.

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

Oettl, A. (2012*a*), 'Honour the helpful', *Nature* **489**(7417), 496–497.

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

Orazbayev, S. (2017), 'International knowledge flows and the administrative barriers to mobility', *Research Policy* **46**(9), 1655–1665.

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

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

Rose, M. E., Georg, C.-P. and Opolot, D. C. (2020), 'Discussants', *Max Planck Institute for Innovation & Competition Research Paper* (20-19).

Rose, M. E. and Kitchin, J. R. (2019), 'pybliometrics: Scriptable bibliometrics using a Python interface to Scopus', *SoftwareX* **10**(July-December), 100263.

Sarsons, H. (2017), 'Recognition for Group Work: Gender Differences in Academia', *American Economic Review* **107**(5), 141–145.

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**(6684), 440–442.

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

A Additional tables and figures

Table A1: Correlation coefficients for the paper sample.

Total citation count													
Top publication	0.51												
# of pages	0.38	0.55											
# of authors	0.15	0.09	0.05										
Auth. total Euclid	0.24	0.24	0.18	0.55									
# of seminars	0.28	0.45	0.38	0.09	0.18								
# of conferences	0.16	0.21	0.21	0.20	0.16	0.44							
# of commenters	0.28	0.41	0.35	-0.03	0.01	0.43	0.34						
Com. total Euclid	0.29	0.46	0.37	0.02	0.17	0.43	0.30	0.77					
Auth. eigenvector (informal)	0.28	0.37	0.27	0.27	0.39	0.29	0.21	0.24	0.35	0.37	0.34		
Auth. betweenness (informal)	0.27	0.31	0.26	0.36	0.54	0.28	0.23	0.22	0.35	0.34	0.32	0.76	
Com. eigenvector (informal)	0.30	0.49	0.37	0.09	0.20	0.40	0.30	0.59	0.68	0.24	0.20	0.52	0.54	
Com. betweenness (informal)	0.28	0.39	0.34	0.07	0.17	0.36	0.27	0.59	0.71	0.20	0.16	0.48	0.54	0.89

Notes: Spearman correlation coefficients for paper sample, where the unit of observation is the a published research paper. See Section 2 for variable definition.

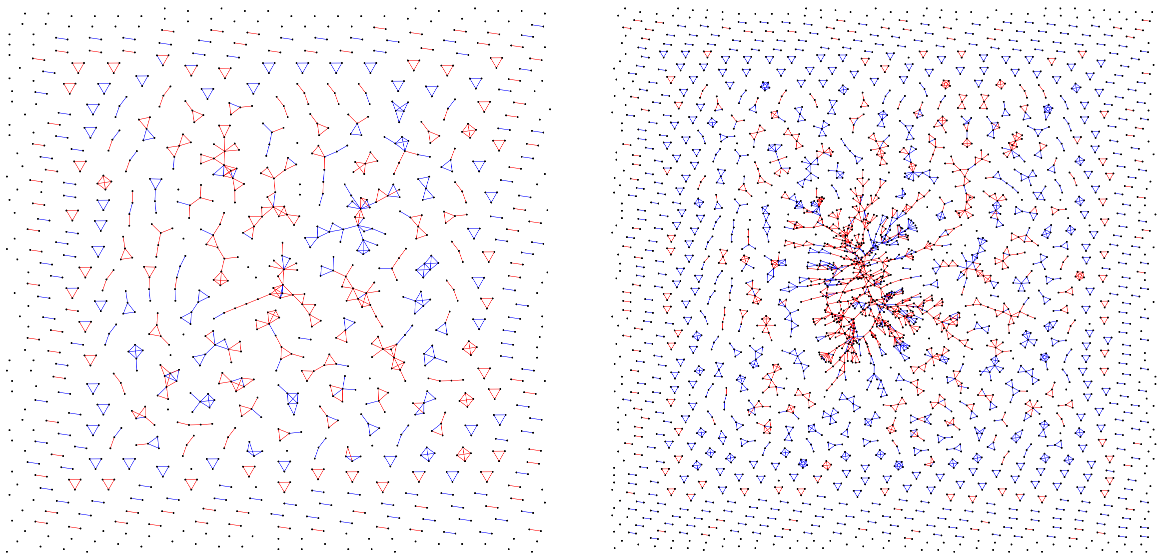
Table A2: Correlation coefficients for the person sample.

Researcher Characteristics											
Euclid. Index											
Publications	0.81										
Citations	0.99	0.87									
Female	-0.13	-0.17	-0.14								
Experience	0.77	0.81	0.80	-0.18							
Network Centralities											
No. of Thanks	0.38	0.22	0.37	-0.09	0.24						
Out-Degree	0.36	0.20	0.35	-0.07	0.23	0.93					
Eigenvector centrality rank (informal)	0.02	0.04	0.03	0.00	0.14	-0.05	-0.03				
Betweenness centrality rank (informal)	-0.09	-0.10	-0.09	0.05	0.02	-0.28	-0.30	0.64			
Degree (co-author)	0.25	0.21	0.26	-0.02	0.15	0.20	0.20	-0.13	-0.20		
Eigenvector centrality rank (co-author)	0.00	-0.06	-0.01	0.06	-0.01	-0.03	-0.02	0.19	0.25	-0.12	
Betweenness centrality rank (co-author)	-0.10	-0.17	-0.12	0.10	-0.06	-0.23	-0.21	0.43	0.51	-0.42	0.60

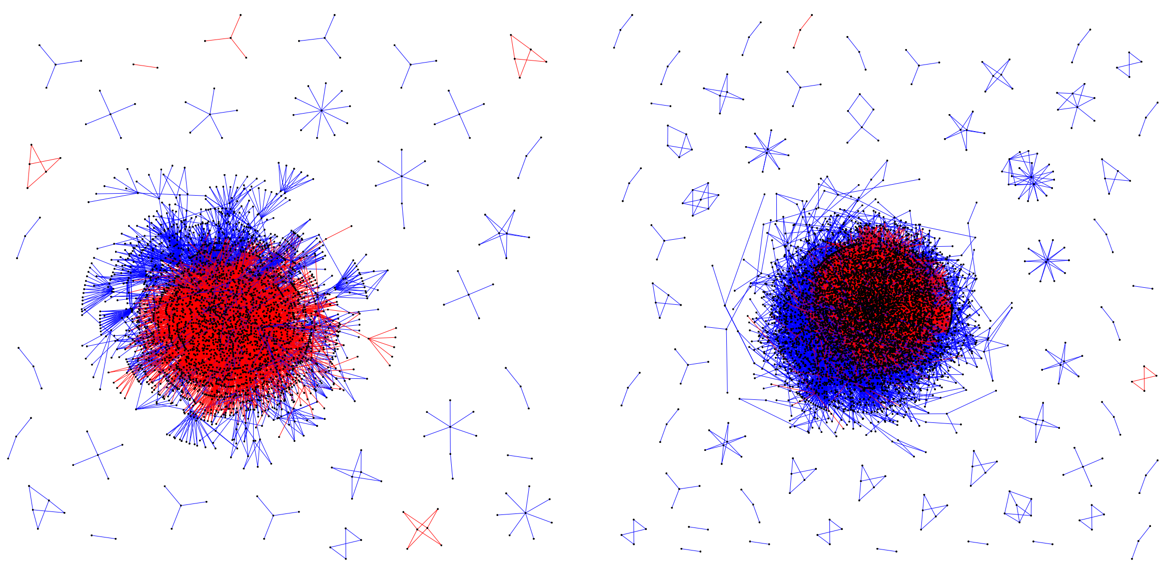
Notes: Spearman correlation coefficients for person sample, where the unit of observation is the combination of researcher i and year t. See Section 2 for variable definition.

Figure A1: 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 networks 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. Graph layout according to the Fruchterman-Reingold algorithm.

B Replication of Existing Studies

B.1 Replication of Ductor et al. (2014)

We follow the methodology of Ductor et al. (2014) as closely as possible.[25] The main point is not to interpret coefficients but to compare the accuracy of a prediction of log future output over the next three years using observables that are available today. The first prediction serves as benchmark where only past output–weighted publication output of the past three years–is used. We add variables to the prediction one group at a time: first variables from the co-author network, then those from the informal collaboration network, finally both combined. Following Ductor et al. (2014), we use the Root Mean Square Error (RMSE) to compare forecast accuracy. A Diebold-Mariano test tests the hypothesis, that a given model and the benchmark model are statistically the same. We use a reduced version of the person sample, namely all those observations that are author-observations. We use the network of informal collaboration and pair it with the corresponding co-author network. Though both networks are not directly comparable to Ductor et al. (2014) due to size and range, they allow comparison with each other.

The variable of interest is productivity. Annual productivity is the weighted publication count of that year. Weights correspond to the SCImago journal rank score factor of the current year or the next available year.[26] Future productivity, the variable to be forecasted, is the log-transformed sum of publications of the next three years. Recent past output, an important predictor that is not always available, is the log-transformed weighted sum of publications of the current the previous four years. Older output, i.e. from start of the career until $t-5$, is subsumed under past output.

Many variables are involved in the forecast. The benchmark contains the cumulative output since the start of a researcher's career until $t-5$, career time dummies, year-dummies, and the number of years since last publication. The author network variables include degree (Equation (6)), degree of order two, membership in the giant component, closeness centrality (Equation (13) in the appendix), and betweenness centrality (Equation (8)). These centralities are all computed in the co-author network of year t. Additionally, the set of variables includes the joint productivity of co-authors as well as of co-authors of co-authors. The commenter network variables include the same variables as before but are exclusively computed in the network of informal collaboration in year t.

[25]There are notable differences to the original study. The authors' dataset covers publications in 850 journals listed in AEA's EconLit database for the 1970–1999 period. Publications are aggregated based on author names only and stem only from these 105 journals. Our data contains publications from 6 journals published in the 1997–2011 period. Publication and productivity includes all publications (indexed in Scopus) however and name disambiguation is more sophisticated.

[26]See www.scimagojr.com. We use the rankings for the "Economics, Econometrics and Finance" and the "Business, Management and Accounting" subject areas. For publications before 1999 we use the impact factor of 1999. Matching rates equals about 20%.

B.2 Replication of Laband and Tollison (2000) and Brown (2005)

To show that our data is akin to the data used in the literature, we replicate the major results of Laband and Tollison (2000) and Brown (2005). Both studies regress a paper's citation count on the amount of informal intellectual collaboration.

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. Controls include the cumulative stock of citations from the previous five years for all authors, as well as the number of pages. The number of commenters is statistically significantly and positively associated with the paper's citation count. In alterations to the model the authors 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, and c) at not belonging to one of the previous groups. Columns (1) through (3) in table A3 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 unlike Laband and Tollison (2000).

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 often 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) of table A3 replicates Brown (2005, Table 8C), with the difference that we do not control for the number of downloads from SSRN. Unlike Brown (2005) 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 A3: Regression results replicating parts of Laband and Tollison (2000) and Brown (2005)

	Six-year citations			Total citations
	OLS			*neg. binomial*
	(1)	(2)	(3)	(4)
Authors' 5-year cites	0.007***	0.006***	0.007***	
	(0.0003)	(0.0003)	(0.0003)	
No. of pages	0.604***	0.659***	0.590***	
	(0.047)	(0.046)	(0.047)	
No. of authors				0.187***
				(0.017)
No. of commenters	0.940***		0.575***	0.021***
	(0.074)		(0.091)	(0.002)
Commenters' 5-year cites		0.001***	0.001***	
		(0.0001)	(0.0001)	
No. of seminars				0.015***
				(0.004)
No. of conferences				−0.003
				(0.008)
Constant	−0.395	2.511**	0.721	5.250***
	(1.276)	(1.254)	(1.281)	(0.079)
Journal fixed effects				✓
Publication year fixed effects				✓
N	5,769	5,769	5,769	5,769
Adjusted R^2	0.148	0.149	0.155	
Akaike Inf. Crit.				64,979.300

Notes: ***, ** and * indicate statistical significance to the 1, 5 and 10 percent level, respectively. Columns (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 percent 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. *Commenters' 5-year cites* is the sum of individual citation stocks for all commenters acknowledged for concept-related input (excluding editors) for the five years prior to the publication year. See Section 2 for definition of other variables.

C Appendix to appear online

C.1 Data collection

A procedure is necessary to link researchers in our dataset with their Scopus Author profiles. For 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 matching 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" or "Business, Management and Accounting", and second it does not include more than 5% of publications in journals outsides these fields. If only one match is found against the Scopus database via a simple name search, we immediately link name and profile. If the search returns fewer 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 a journal where no one else in our database published in, and if necessary correct manually.

C.2 Strategic Acknowledging

An important question to address is whether acknowledgements contain unbiased information at all. While the different parts of an acknowledgement section undoubtedly come with different degrees of doubt regarding their validity, the names of acknowledged researchers are typically the most contested part. Since we are interested mostly in the names of acknowledged researchers, which we claim to reveal informal collaboration, the biggest thread comes from “strategic acknowledging”. We define two forms of strategic acknowledging. The first follows Hamermesh (1992, p. 171) and describes the author’s attempt to influence an editor by acknowledging "someone who has not seen the paper, as a talisman against that person being chosen" as referee. Even though there might be conflicting views, the general assumption seems to be that editors do not pick already acknowledged commenters.[27] According to this assumption, authors would want to thank someone that has a reputation of being a tough referee ("Cite your friends, acknowledge your foes."). In the second form of strategic acknowledging authors try to signal quality to readers, most importantly to the editor in order to increase publication chances.

If these forms of strategic acknowledging exist, they can harm our analysis in two ways: First,

[27] Other authors have indicated they believe that editors would prefer to pick acknowledged commenters. However, editors of various journals 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.

authors acknowledge researchers they didn't actually speak to but we nonetheless assume a link between those two. Second, authors actually spoke to the researcher but only in order to acknowledge her, while what the person recommended or suggested in order to improve the paper is secondary. We believe authors do not usually acknowledge someone who has not actually given a comment. This would be outright fraud and carries a high reputational risk. If that person learns about it (e.g. during the review process), it will reflect very badly on the author. In his "Guide to Professional Etiquette" Hamermesh (1992, p. 171) writes "DON'T PLAY THESE GAMES - the gains are not worth the potential cost of being caught" (emphasis in the original). That authors approach seniors and more well-known researchers in the field is certainly prevalent, but we believe it is rare that authors do not make use of a comment or suggestion from that person. 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.

Nonetheless, we do not observe that authors predominantly acknowledge researchers with high reputational value. We observe that half of all papers acknowledge individuals that no other publication in our dataset acknowledges. Also, as Section 5.4 shows, being thanked often and being prolific (i.e., having a high reputational value) correlate only mildly (Figure 8). Finally, the vast majority of authors sort acknowledged authors alphabetically, and not by signalling value. These observations speak 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.

Finally, both forms of strategic acknowledging predict similar behaviour of authors which we do not observe in our data. In order to increase the strength of the signal, authors would sort acknowledged commenters by signalling content to avoid that an influential name gets lost. Instead, we observe that commenters are sorted alphabetically. If there is ordering, then to favor editors and anonymous referees. A similar argument applies to the order of the various forms of informal collaboration: We do not observe that commenters are always listed first in the acknowledgement section. Sometimes authors list seminars, conferences, research assistance or funding before commenters.

C.3 Closeness centrality

Closeness centrality is 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 researcher i to other researchers:

$$\text{closeness}_i = \frac{n-1}{\sum_{j \neq i} \sigma(i,j)} \tag{13}$$