Discussants

Article *in* SSRN Electronic Journal · January 2020

DOI: 10.2139/ssrn.3727173

Daniel Christopher Opolot
University of Cape Town

Discussants*

Michael E. Rose[†1,2], Co-Pierre Georg[3], Daniel C. Opolot[3], and [2]

[1]Max Planck Institute for Innovation and Competition

[2]Kiel Institute for the World Economy

[3]University of Cape Town

2

Abstract

We study the role of informal collaboration in academic knowledge production. Our focus is on published papers presented at workshops at ten NBER Summer Institutes. Though not randomly selected, papers are of comparable quality pre-discussion and workshops are similar. Even among this set of papers that is highly selected on expected quality, discussants matter for top journal publication. Conditional on having a discussant, a paper's citation count increases in the discussant's prolificness. This hints at quality-improving channels. Conversely, using social network analysis we rule out a diffusion channel through which citations accumulate because discussants diffuse information about the paper.

Keywords: Informal collaboration, academic production function, NBER Summer Institutes

JEL Classification: A14, D83, 031, O33

*Rose: Max Planck Institute for Innovation and Competition and Kiel Institute for the World Economy (email: michael.rose@ip.mpg.de); Georg: University of Cape Town (email: co-pierre.georg@uct.ac.za); Opolot: University of Cape Town (email: daniel.opolot@uct.ac.za). We wish to thank Pierre Azoulay, Fabian Gaessler, Dietmar Harhoff, Adam Jaffe, Danielle Li, Atif Mian, Paula Stephan (discussant), Rainer Widmann and especially Jeff Furman. We are also grateful to seminar participants at Maastricht University, University of Cape Town, Stellenbosch University, the Kiel Institute for the World Economy and KU Leuven as well as to conference participants at the 2016 Africa meeting of the Econometric Society, the INET-YSI "Innovation, Economic Complexity and Economic Geography" Workshop at MIT, the 13th Workshop on the Organisation, Economics and Policy of Scientific Research at the University of Bordeaux, the Academy of Management 2019 Annual Meeting, the 2019 Annual Meeting of the European Economic Association, and the 2019 meeting of the Verein für Socialpolitik for helpful comments and feedback. Carolin Formella and Nurzhan Sapargali provided valuable research assistance.

†Corresponding Author. E-Mail:

1 Introduction

Informal collaboration—receiving feedback and commentary from colleagues on ongoing research papers—is commonplace in economic research. More informal collaboration, measured as the number of acknowledged commenters, seminars and conferences, is associated with a higher likelihood that a paper is accepted for publication and a greater number of citations once it is published (Laband and Tollison, 2000; Brown, 2005; Rose and Georg, 2018).[1] Through which channel this correlation arises is, however, not well understood. This is surprising given the amount of time and resources informal collaboration activities (e.g. attending and organizing conferences, and private correspondence among colleagues) consume.[2]

We study the role that discussants—a special type of informal collaborators—play in academic knowledge production and the dissemination process. Our laboratory is the assignment of discussants at the National Bureau for Economic Research's Summer Institutes (NBER SIs). We find that discussants matter for the publication process, even among a set of papers that is already highly selected on expected quality. A published paper previously discussed at an NBER SI is more likely to be published in a prestigious journal than those without. We also find that, conditional on having a discussant, a one standard deviation increase in the discussant's prolificness corresponds to a 9.7% increase in citation count for the average paper, corresponding to 13 more citations over the lifetime of a paper.[3] Our findings are consistent with the existence of various *quality channels* whereby having a discussant improves the inherent quality of the paper. Conversely, we do not find evidence for the existence of a *diffusion channel* whereby discussants may act as seeds of diffusion of information about the paper.

The NBER SIs are a high-profile annual three-week series of workshops that showcase the latest research advances across subfields in economics. Each SI hosts a range of different workshops corresponding to topical groups and programs. We focus on the workshops of ten Finance-related groups and programs. Workshops have one of two alternative organizational structures

[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] According to the academic portal `https://www.econbiz.de/Events/Results`, there were at least 900 conferences in economics and business in 2019. There are about 100,000 meetings in medical science each year (Ioannidis, 2012).

[3] We regard this estimate as lower bound, as it is not necessarily the discussant's object function to improve the paper, rather than leaving a good impression with the audience.

that we exploit: Either they always feature discussants (usually together with a discussion with the audience), or they never feature discussants (i.e. where presenters have a discussion with the audience only).[4]

The effect of a presentation plus having a discussant (and potentially a general discussion) versus a presentation plus having a general discussion only is well identified if (i) neither authors nor organizers sort into workshops based on the fact that they have discussants; (ii) papers are of comparable quality; and (iii) the workshops are otherwise identical. We provide evidence that all three assumptions hold. First, we have conducted a survey among organizers of the NBER SIs. The qualitative evidence from the survey suggests that authors do not decide which NBER workshop to submit to based on whether it features discussants. We also do not find any indication of differences in the quality of accepted papers between the two categories of workshops. The NBER SIs are all highly competitive and prestigious, and hence, it is much more important for authors to achieve a good topical fit than have a discussant. Second, we examine the characteristics of presentations and discussants. We find that presentations without discussants are on average as long as presentations with discussants; and that published manuscripts in our sample cite similar journals irrespective of which category of workshop (with or without discussant) they were presented in.

Despite the above evidence, the identification of the effect of a discussant on a paper's success may be hampered by unobserved heterogeneity across NBER workshops. The approach that is typically used to alleviate unobserved group heterogeneity is to introduce group fixed effects. However, we cannot introduce group fixed effects because whether or not discussants are part of a workshop correlates perfectly with the NBER working group. We adopt two approaches to address the potential unobserved group heterogeneity. First, in an alternative specification, we account for topics to control for the effect of topical differences across groups. Second, in yet another specification, we select only the topically most similar groups. We find that our estimates are robust to these alternative specifications, supporting the assumption that workshops are identical in aspects other than whether or not they feature a discussant.

[4] The reasons for this organizational difference are lost in history, but the differences are persistent. From the short survey we conducted among organizers, one respondent speculated that the difference might be due to "a different norm: everyone reads the papers beforehand, so everyone forms an opinion and is able to ask informed questions," while another called it just "path dependence."

On the other hand, the effect of a discussant's characteristics on a paper's academic success is well identified if (i) organizers match discussants to papers and authors based primarily on topical fit (e.g. whether the discussant knows the relevant literature, methods and data) rather than discussant characteristics (e.g. prolificness, age and gender); and (ii) discussants discuss a paper irrespective of which session it is in. Although discussants are not assigned randomly, our survey among the NBER SI organizers in our sample shows that they do not match paper and discussants based on experience or prolificness.[5] As survey respondents confirmed that discussants usually accept, we rule out sorting by discussants. Such sorting would arise if potential discussants decline to discuss a paper. Being a discussant at a NBER SI is considered a prestigious opportunity and a signal of high standing within the profession.

In our survey we ask three questions: "What do you look for when accepting papers, and how do you match discussants to papers? How exogenous is the process to authors?" With regards to the first question about the paper selection process, the general consensus is to look for a "mix of quality, novelty and fit" with particular emphasis on novelty. Some organizers also mention the importance of an author's ability to present well, which is even more relevant for the selection of discussants. Indeed, almost all organizers highlight how important it is for them to have a discussant that can provide "a basis for a lively, productive debate between authors, discussants, and the audience." Furthermore, almost all organizers highlight the importance of a topical fit between authors and discussant, mentioning that discussants are "often authors of good papers that were not chosen for presentation." Some respondents also emphasize that discussants should not be "too close to the author and if possible coming from a different perspective" and one even highlights that discussants should have "No fear of authors (i.e. probably dont get a very junior person to discuss a big shot, unless you know the junior person is fearless). With regards to our third question about the exogeneity of the discussant selection process for authors, few organizers explicitly answered the question and those who did emphasized that they "didn't consult with authors." Consequently, we do not find assortative matching in the observables in our sample: Seniority or prolificness of authors of discussed manuscripts

[5]There are 66 different organizers in our sample. We contacted 60 of them for which we found valid e-mail addresses; 15 responded but 4 respondents said they were not engaged with the selection of discussants. Our knowledge thus relies on the insights of the remaining 11 respondents, who were engaged in the organization of 42 distinct workshops.

are not correlated with seniority or prolificness of discussants.

However, a confounding factor possibly arises from the ability of the organizers to attract discussants. We thus introduce fixed effects for the corresponding NBER working group, and cluster standard errors on this level as well. The group-fixed effects also capture the organizer effects since organizers seldom change during our sample period.

Our samples are constructed from presentations in Finance-related workshops at NBER SIs between 2000 and 2009.[6] We focus on Finance as a subfield to rule out subfield-specific confounders. In total, 922 presentations took place at a total of 85 workshops in this period. Of these, 696 (75%) resulted in publication by January 2020. Of the 696 publications, 50% got published in the top 3 Finance or top 5 Economics journals, and 50% got published between two and five years after presentation.

We measure academic success of published papers in three ways: the total citation count by January 2020 according to Scopus; whether the publishing journal is one of the commonly denoted "top" journals in Financial Economics; and the Journal Impact Factor.Despite the shortcomings these measures may have, it is reasonable to expect that they are correlated with inherent paper quality and are measures that many authors care about. For example, the journal a paper is published in is an important metric to evaluate academic economists in tenure decisions at most universities.

Our results can be explained by different quality channels through which a discussant might affect a paper's academic success. First, discussants may improve papers directly via helpful comments. Since discussants at NBER SIs put in a lot of effort and are much more elaborate than in other conferences, a discussion at the NBER SI is a strong treatment. Colleagues who presented at a NBER SI spoke of the feeling that "the paper was re-written" by the discussant. As such, we think of a discussant as a free referee. And second, discussants might motivate authors to work harder on their papers (in the pre- or post-discussion period), link authors to other helpers in the community and inspire thoughtful comments from workshop attendees. Our analysis is not able to distinguish between these two channels. If this channel exists, we would expect the discussion's impact to depend on the characteristics of the discussant. For

[6]The lower bound for this range is due to the availability of source data and is upper bounded due to limitations of data available for robustness checks.

example, a knowledgeable (i.e. experienced) discussant has the expertise on how to structure papers for top journal publication. Thus, a discussion by a more prolific discussant should significantly improve the quality and in turn the academic success of a paper.

Finally, discussants might improve the academic success of a paper without improving the quality, namely by an attention effect (hearing the same thing twice, which is therefore more memorable), by promotion to potential referees, or by dissemination to other authors in the field so that they cite the paper. In the second part of the analysis, we focus on discussants' characteristics to examine how they correlate with academic knowledge production and dissemination.

In the last part of our analysis we specifically probe the diffusion channel. Discussants, being among the first to learn about the existence and more importantly the quality of the paper, may become seeds of diffusion outside of the SI. That is, discussants may start a diffusion process where they tell their colleagues about the paper, who then tell their colleagues, and so on.[7] To estimate a discussant's diffusion ability directly, we propose a new network centrality measure—neighborhood centrality.[8] For each discussant in the network, the neighborhood centrality counts the number of all nodes within a given distance from a researcher, while discounting distant connections. We test neighborhood centrality in two different social networks. One network captures co-authorship relationships in Finance and related fields of up to about 50k authors. The other network captures fine-grained informal collaboration links of up to about 7.5k authors within Finance (Rose and Georg, 2018). We do not find any evidence of a diffusion channel in either networks.

Our findings contribute to the empirical work studying the effect of informal collaboration on academic success of papers (Laband and Tollison, 2000; Brown, 2005; Castaldi et al., 2015; de Leon and McQuillin, 2020; Gorodnichenko et al., 2019). However, except for de Leon and McQuillin (2020), these papers can only study correlations. de Leon and McQuillin (2020) are the first to use quasi-experimental evidence, namely, the sudden cancellation of the 2012 American

[7] This form of network diffusion is akin to the diffusion of information about products, practices and services (Bandiera and Rasul, 2006; Conley and Udry, 2010; Banerjee et al., 2013). These papers show that information sharing in networks of friendship, peer-to-peer, colleagues and family relations, influence individual decisions, and hence, successful diffusion of products, services and practices.

[8] As we will discuss later, several other centrality measures developed in the literature are not well suited to the form of diffusion we wish to examine.

Political Science Association Annual Meeting. They estimate the positive impact of conferences on papers' citation counts. Our study provides the first attempt at causal evidence, although not perfect, of the benefits of having discussants at conferences. This literature stream is embedded in a broader literature on peer effects in knowledge production and diffusion. Prominent examples include Waldinger (2012) who finds positive peer effects among members of academic departments; Azoulay et al. (2010), Oettl (2012*b*) and Borjas and Doran (2015) who report evidence of knowledge spillovers among former co-authors.

Our paper is also related to the literature that seeks to understand information diffusion within scientific communities. de Leon and McQuillin (2020) find that conferences act as avenues for advertising papers through face-to-face interactions. Similar effects were documented by Belenzon and Schankerman (2013) for inventors. Recently, Baruffaldi and Pöge (2020) show that diffusion of scientific insights to industry is greatly facilitated by conferences. We find no evidence of diffusion outside of conferences, which may suggest that the diffusion of papers through online publication is more prominent than through face-to-face interactions. However, we only consider the diffusion effects initiated by discussants. It is possible that when all conference session participants are considered, the diffusion effects will become significant.

Overall, our results are in line with recent empirical evidence suggesting that informal collaborators deserve more credit with regards to their contributions to paper authorship than they currently receive (Ponomariov and Boardman, 2016; Oettl, 2012*a*). Our findings also lend credence to the idea that not only do academic conferences play a role in knowledge production and dissemination, the manner in which they are organized affects how strong their impact is.

2 Variable construction

2.1 NBER Summer Institutes

We study the effects of informal collaboration on a paper's academic success using presentations in Finance-related workshops at the NBER SIs. NBER SIs are annual meetings with distinct workshops by NBER working groups or programs.[9] Some of the workshops always feature

[9] We do not distinguish between groups and programs, as their workshops follow the same procedure.

discussions by discussants specifically invited for their discussion. From the program of every NBER SI since 2000, we source the title of each paper and the names of authors and discussants, and for each author, the affiliation.[10] We focus on the SIs between 2000 and 2009 of the following NBER working groups: "Asset Marketing/Real Estate" (AMRE), "Asset Pricing" (AP), "Corporate Finance" (CF), "Impulse and Propagation Mechanisms" (EFCE), "Capital Markets in the Economy" (EFEL), "Forecasting and Empirical Methods in Macro and Finance" (EFFE), "International Finance and Macroeconomics" (IFM), "Monetary Economics" (ME), "Finance and Macro" (MEFM), "Economics of Real Estate and Local Public Finance" (PERE) and "Risk of Financial Institutions" (RISK).[11]

In total, 922 presentations took place at 85 distinct workshops between 2000 and 2009. This includes 36 presentations where the same papers were presented at least twice, and 4 presentations whose title is not mentioned in the program.[12] However, not every presentation eventually resulted in a publication:[13] A total of 696 (75%) of the presentations resulted in publication by January 2020.

Table 1 gives an overview of the number of presentations by year and NBER group. The share of papers published by group ranges from 63% for PERE working group to 86% for the AP working group. Three workshops, denoted in italics, were jointly held by two groups: IFM 2001 (4 presentations joint with EFEL), RISK 2007 (completely joint with CF), RISK 2008 (completely joint with AP). In the analysis we include fixed effects for both groups if appropriate and cluster on the joint group.

Among the 696 publications are 12 publications in or as books, 4 policy publications and 21 Society of Economic Dynamics proceedings. The rest got published in 86 different journals. The most important outlets for the published papers are *The Journal of Finance* (62 papers) and *The American Economic Review* (60). They are followed by *The Review of Financial Studies* (49), the *Journal of Financial Economics* (44) and the *Journal of Monetary Economics* (42).

[10] See `http://www.nber.org/summer-institute/`.

[11] Some groups change abbreviations through time. For better comparison, we use the abbreviations as of 2020, or the last available year.

[12] When a manuscript is presented twice but with a different title, we are unable to map them.

[13] Many papers change their titles. We, therefore, conducted an extensive and thorough internet search for each paper based on the authors and abstracts to identify those papers with a different title.

Table 1: Presentations in NBER SIs of financial NBER working groups, by group or program and year.

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	total	share
AMRE		5 (6)	5 (6)								10 (12)	83%
AP	8 (8)	5 (8)	8 (9)	8 (10)	6 (9)	9 (9)	9 (9)	7 (9)	11 (12)	8 (8)	79 (91)	87%
CF	9 (10)	4 (7)	7 (9)	9 (11)	10 (13)	13 (14)	10 (11)	11 (13)	16 (18)	11 (12)	100 (118)	85%
EFCE	11 (15)	7 (17)	8 (17)	7 (15)	14 (17)	10 (15)	13 (15)	11 (15)	10 (15)	9 (15)	100 (156)	64%
EFEL		13 (16)	13 (15)	11 (12)	6 (11)	9 (11)	11 (12)	10 (12)	8 (10)	9 (12)	90 (111)	81%
EFFE	11 (14)	10 (12)	9 (12)	8 (12)	10 (12)	9 (12)	7 (12)	9 (12)	6 (12)	8 (12)	87 (122)	71%
IFM	6 (8)	7 (8)*	11 (12)	8 (11)	10 (12)	5 (8)	8 (10)	7 (10)	9 (12)	11 (14)	82 (105)	78%
ME	8 (10)	5 (8)	7 (11)	8 (9)	9 (12)	7 (12)	13 (14)	7 (12)	9 (13)	12 (13)	85 (114)	75%
PERE	4 (5)	6 (6)	4 (6)	1 (6)	6 (9)	7 (10)	5 (10)	7 (10)	4 (11)	8 (9)	52 (82)	63%
RISK							3 (3)	*1 (2)*	*2 (3)*	11 (12)	17 (20)	85%
total	57 (70)	59 (84)	72 (97)	60 (86)	71 (95)	69 (91)	79 (96)	69 (93)	73 (103)	87 (107)	696 (922)	
share	81%	70%	74%	70%	75%	76%	82%	74%	71%	81%	75%	

Notes: This table lists the total number of presentations and the number of presentations that resulted in publication. The NBER groups and programs are Asset Marketing/Real Estate (AMRE), Asset Pricing (AP), Corporate Finance (CF), Impulse and Propagation Mechanisms (EFCE), Capital Markets in the Economy (EFEL), Forecasting and Empirical Methods in Macro and Finance (EFFE), International Finance and Macroeconomics (IFM), Monetary Economics (ME), Finance and Macro (MEFM), Economics of Real Estate and Local Public Finance (PERE) and Risk of Financial Institutions (RISK). Groups or programs printed in bold include discussants. The first number is the number of presentations that resulted in a Scopus-indexed publication by January 2020, the number in brackets is the total number of presentations. Values in italics indicate that the respective workshop was jointly held with another working group. Rows "total" and "share" account for joint sessions. Includes 4 presentations without known title.

2.2 Paper characteristics

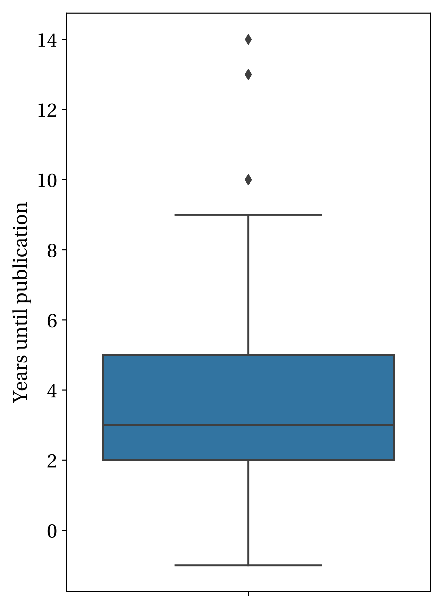
For each of the papers presented at the NBER SIs that got published, and for which bibliometric information is available, we compute three measures of academic success.The first measure is the total count of citations a paper garnered by January 2020. Second, we use a binary variable equal to 1 indicating publication in one of the top three Finance journals or top five Economics journals, and 0 otherwise.[14] Finally, we use the SCImago Journal Impact Factor of the journal from the year before publication. This iterative impact factor uses the Eigenscore method that weights citations to individual documents of the past three years by the Journal Impact Factor

[14]These journals are *Journal of Financial Economics*, *Review of Financial Studies*, and *The Journal of Finance*, as well as *Econometrica*, *Journal of Political Economy*, *The American Economic Review*, *Quarterly Journal of Economic* and *Review of Economic Studies*.

of the citing articles. [15]

Although it is reasonable to expect that these measures correlate with true paper quality, none of them is without criticism. First, the review process, with all its known inefficiencies, and the focus on few journals as the quality-setting standard (Heckman and Moktan, 2020), implies that journal status is but a noisy signal of paper quality. Citations by other academic papers partly depend on factors that are not related to quality at all. For example, Coupé et al. (2010) and Feenberg et al. (2017) show that citation counts are subject to the ordering in the issue; and the "Matthew effect" (Merton, 1968) (i.e. a phenomenon where high citation counts beget more citations) has been observed in all disciplines. Second, since editors and referees might err occasionally (as Gans and Shepherd (1994) document) some papers published in lower-ranked journals regularly receive more citations than papers published in top journals (Oswald, 2010). Despite these shortcomings, our measures of academic success are what many academic actors care about. For example, the journal a paper is published in is arguably the most important metric to evaluate academic economists in tenure decisions.

Figure 1: Distribution of number of years between presentation and publication.



Notes: This boxplot shows the distribution of the publication lag, i.e. the number of years between the presentation and the publication year.

[15]The impact factors can be found at http://www.scimagojr.com/journalrank.php. SCImago uses the Scopus data. As alternative measures of journal quality, we use the journal's h-index and the average citation count over the previous five years. Results are qualitatively the same.

Finally, we obtain the number of pages of the published version of the article, and count the number of years between the presentation at an NBER SI and the final publication. We call the latter variable *age*. It has two purposes: First it is an (imperfect) indicator of how much the article changed subsequent to the presentation, but is also sheds some light on publication lags. Figure 1 shows that this publication lag can increase to as much as 13 years. The average manuscript gets published in the third year following presentation (conditional on getting published by January 2020); 50% of the presented manuscripts get published between 2 and 5 years after presentation. The page count and the number of citations were retrieved in January 2020 from Scopus using the code developed by Rose and Kitchin (2019).

2.3 Author and discussant characteristics

For each of the 257 discussants and the 919 authors of the papers in the journal and discussants samples, we compute three variables. We denote them as author and discussant characteristics, respectively. Author characteristics are computed in the year before publication; discussant characteristics are computed in the year of the discussion. Otherwise, variable definition is the same. We compute them from Scopus data as of January 2020 using the code developed by Rose and Kitchin (2019). Unlike some other popular bibliometric databases, Scopus features unique author profiles so that the researcher does not have to disaggregate author names.

We measure prolificness using the Euclidean index of citations (Perry and Reny, 2016). Its computation algorithm is as follows: For each year t, count the total number of citations to each of researcher i's n publications published by t, then take the square root of the sum of the squared citation counts. That is, if $c_{i,k,t}$ is author i's total citation count for paper k by period t, then $\text{Euclid}_{i,t}$ of author i at period t is:

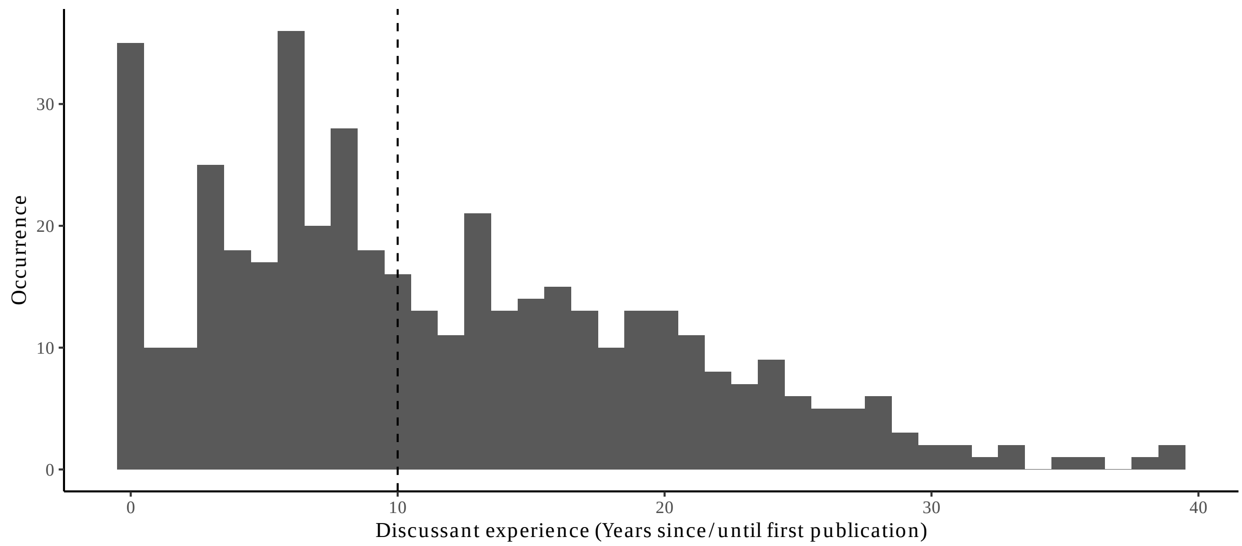
$$\text{Euclid}_{i,t} = \sqrt{\sum_{k=1}^{n} c_{i,k,t}^2} \tag{1}$$

A researcher's Euclidean index increases monotonically in the number of publications with positive citation count.[16] As an example, consider an author with a stock of two publications

[16] Alternative measures would include the total number of publications or the total citation count normalized by the number of years of experience. The Euclidean index however takes into account both measures making it a more accurate measure.

with 5 and 50 citations by t. The Euclidean index of citations equals $\sqrt{5^2+50^2} \approx 50.25$. If the first paper's citation stock increases from 5 to 10 in $t+1$, and the author publishes a third paper that garners 2 citations, the Euclidean index of citations increases to $\sqrt{10^2+50^2+2^2} \approx 51.03$.

Figure 2: Histogram for a discussant's academic experience in year of discussion.



Notes: This figure shows the histogram of a discussant's experience in the year of the discussion. Experience is the number of years since the first recorded publication in Scopus, or 0 if the first publication is in the future. The dashed line indicates the median value (10 years).

The second characteristic is "experience," which for any year t is the number of years since the researcher's first publication. If the first publication is in the future, as may happen for some discussants, the experience is 0. The median experience of a discussant at the time of discussion is 10 years (Figure 2). About 7% of all discussions were delivered by discussants before or in the year of their first publication (i.e. where discussants have an experience value of 0). These include mostly junior academics but also practitioners who follow different publication strategies.

2.4 Co-author and informal collaboration networks

We study the diffusion of information about new papers through interactions that occur during academic collaboration. The two forms of collaboration we study are co-authorship and informal intellectual collaboration (Laband and Tollison, 2000). To capture these interactions, we construct a co-author and an informal collaboration network where nodes are researchers and links between them are co-authorship and informal intellectual collaboration, respectively.

We construct the co-author network from research articles, reviews and conference proceedings published in 360 journals. We include a journal if it is ranked "C" or higher in selected journal rankings of Combes and Linnemer (2010). We include every ranking by JEL category with at least three publications in our sample. These are: Finance (JEL category G), Micro/Game Theory (D), Public/Political Science (H), Law and Economics (K), Macro/Monetary (E), International (F), Urban/Regional (R) and Econometrics (C).

We use the publications from these 360 journals to construct the co-author networks. Let $\mathscr{A}_t$ be the set of papers published in years $\{t, t+1, t+2\}$ and $\mathscr{N}_t$ be the corresponding set of authors of these papers. The co-author network of year t is a graph, $\mathscr{G}(\mathscr{N}_t, \mathscr{E}_t)$, consisting of $\mathscr{N}_t$ authors of the $\mathscr{A}_t$ papers and a set $\mathscr{E}_t$ of undirected links connecting the $\mathscr{N}_t$ authors. An undirected link between any pair of authors $i, j \in \mathscr{N}_t$ exists if they have co-authored at least one paper in the set $\mathscr{A}_t$. To study an information diffusion channel, we consider forward-looking networks to capture the diffusion of information about the existence and quality of papers presented at the NBER SIs in year t through co-authorship interactions.

The informal collaboration network is constructed using the CoFE (Collaboration in Financial Economics) dataset of Rose and Georg (2018). The dataset contains manually collected acknowledgements to individual researchers from 5,759 research papers published between 1997 and 2011 in six Finance journals with similar topical focus.[17] Informal collaborators include all persons that are acknowledged intellectual input. We exclude referees and the managing editor(s) for the year of publication and the previous two years. This prevents overestimating the editors' position in the network.[18] Rose and Georg (2018) also provide evidence that strategic concerns are not the driving motive behind an acknowledgement. This is crucial for the usage of this data as representation of actual knowledge flow.

The network of informal collaboration is constructed similarly, with three small differences. First, $\mathscr{A}_t^i$ is the subset of 5,759 research papers that were published in years $\{t, t+1, t+2\}$, where

[17] *The Journal of Finance*, the *Journal of Financial Economics*, *The Review of Financial Studies*, the *Journal of Financial Intermediation*, the *Journal of Money, Credit, and Banking* and the *Journal of Banking and Finance*.

[18] 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, it will to be biased towards journals with a higher publishing frequency.

t ranges from 1997 to 2009. Second, the set of nodes consists of all authors and commenters acknowledged on any of the papers. And third, links are directed author-to-commenter relationships. Due to a much smaller set of journals, the networks of informal collaboration are much smaller than the co-authorship networks. However, both networks grew significantly between 2000 and 2009. The size of the co-authorship network and network of informal intellectual collaboration increased by over 50% from 27,712 and more than doubled from 3,722 researchers, respectively.

2.5 Information diffusion in social networks

One possible channel through which discussants could affect a paper's citation count is by simply telling their contacts (e.g. co-authors, commenters) about it, who would then be more likely to learn about the paper. Existing measures of network centrality are not well suited to measure the flow of information in our case, however. First, we cannot simply use Eigenvector centrality, which is a measure of influence in models of opinion formation through averaging (Golub and Jackson, 2010). Its generalized variants, Katz-Bonacich centrality and diffusion centrality, capture node influence in strategic interactions (Ballester et al., 2006; Banerjee et al., 2013). Diffusion centrality, which nests Katz-Bonacich centrality, is expressed as the discounted sum of powers of the network's adjacency matrix, and Banerjee et al. (2013) interpret it as the *expected number* of times all agents hear about the information. All three measures include repeated interactions, which would be equivalent of a discussant relaying information about a paper to her colleagues, who then relay it back to her.[19] Instead, we expect that learning about an existing paper *once* is sufficient for researchers to cite the paper.

We introduce *neighborhood centrality*, a novel centrality measure, to estimate a discussant's ability to diffuse knowledge about a paper. The diffusion process we model starts from the discussant and propagates outwards to her colleagues, and to colleagues' colleagues, and so on, with potentially diminishing effects.

Formally, let $k_{i\tau}$ be the number of τ-th order neighbors of i. That is, $k_{i\tau}$ is the number of all

[19] In addition, the interpretation of diffusion centrality adopted by Banerjee et al. (2013) assumes that if an agent receives information from k different sources in period $t-1$, she must transmit that information independently k times to her neighbors in period t. Relaxing this assumption—which is not likely to hold in our case—already leads to a different notion of centrality (Genicot and Bramoullé, 2018).

nodes at distance (measured as the number of links in network $\mathcal{G}$) τ from i. Let δ be a discount factor of information decay. Then, the neighborhood centrality of i is defined as:

$$n_i(\delta) = \sum_{\tau=1}^{\infty} \delta^{\tau} k_{i\tau} \tag{2}$$

Appendix B gives examples and details a simple algorithm to compute neighborhood centrality. The relevance of i's information to her distant neighbors is discounted by $0 \leq \delta \leq 1$ to account for topical and expertise distances.

When δ is small, for example, when there is little topical overlap between researchers, neighborhood centrality will be strongly correlated to node degree. The closer δ is to one, the more uniform neighborhood centrality across nodes becomes. We do not take a stance on what the "correct" value for delta is and instead report results for a range of information decay parameters.

For technical reasons, neighborhood centrality only uses the largest component of the network, defined as the largest group of nodes that are connected via intermediate sequences of nodes and links.

2.6 Samples and summary statistics

We use three different samples throughout the analysis, summarized in Table 2. The "Presentations sample" consists of all manuscripts with known titles presented once at Finance-related NBER SIs between 2000 and 2009. The sample includes 845 observations. We use the Presentations sample to descriptively understand the relationship between publication probability and discussion. In this sample, 60% of manuscripts had a discussant.

The "Journal sample" consists of all papers from the Presentation sample that (a) were published in any peer-reviewed journal by January 2020, (b) has bibliometric information available (this excludes 6 observations)[20] and (c) were discussed at most once. The sample includes 592 observations, 385(65%) of which have had a discussant. We use the Journal sample to under-

[20]We use data provided by Elsevier Scopus. Scopus does not index every journal completely; some volumes for some non-Elsevier journals are missing, even if the entire journal is covered. This is idiosyncratic and not due to editorial reasons: When the database was initially populated, Scopus simply did not receive meta information in sufficient quality. There is, for instance, one publication in the *B.E. Journal of Macroeconomics* that Scopus doesn't index before 2007.

stand the difference between having a discussant and not having a discussant (and instead a possibly longer general discussion with the audience). The average paper received 128.4 citations since publication and got published in a journal whose Journal Impact Factor equals 8.0.[21] 51.4% of the observations got published in a top five Economics or top three Finance journal. Table A1 shows that top publication status and Journal Impact Factor correlate highly with each other, but not highly with citation count. This suggests that journal quality and citation count capture distinct aspects of academic knowledge production and dissemination. None of the three measures correlate highly with the discussant indicator.

The "Discussants sample" includes 346 papers from the Journal sample where we know the identity of the discussant from the program (for 4 discussed manuscripts this is not the case). This sample serves to understand the effect of a discussant's characteristics. Apart from the smaller size (346 observations instead of 592), the sample is very similar. The average paper received 135.9 citations, the impact factor of the publishing journal equals 9.0 on average, and 63% of the papers got published in a top Economics or top Finance journal. The average discussant has 11.6 years of experience in academic publishing. We document low correlation coefficients between author and discussant characteristics in Table A2.

[21] To put this number into perspective: In 2009 *The American Economic Review*'s Journal Impact Factor equalled 8.1. The highest Journal Impact Factor equals 27.2 and characterizes the *Quarterly Journal of Economics* in 2014.

Table 2: Samples and their summary statistics.

(a) PANEL A: Presentations sample

	N	Mean	Median	Std.Dev.	Min	Max
Published	846	0.7	1	0.43	0	1
Has discussion	846	0.6	1	0.49	0	1

(b) PANEL B: Journal sample

	N	Mean	Median	Std.Dev.	Min	Max
Total citations	592	128.4	72	178.24	0	1519
Top publication	592	0.5	1	0.50	0	1
Journal Impact Factor	571	8.0	7	5.13	0	23
# of pages	592	29.6	29	11.55	5	87
# of authors	592	2.2	2	0.80	1	5
Age	592	3.5	3	2.06	−1	14
Author total Euclid	592	427.8	213	700.80	0	8114
Author total experience	592	23.5	22	16.94	0	89
Has discussion	592	0.6	1	0.48	0	1

(c) PANEL C: Discussants sample

	N	Mean	Median	Std.Dev.	Min	Max
Total citations	346	135.9	84	183.81	0	1519
Top publication	346	0.6	1	0.49	0	1
Journal Impact Factor	339	9.0	9	5.37	0	23
# of pages	346	30.8	31	10.99	5	65
# of authors	346	2.3	2	0.80	1	5
Age	346	3.7	3	2.22	−1	14
Author total Euclid	346	425.5	216	718.97	0	8114
Author total experience	346	23.1	21	16.99	0	89
Discussant Euclid	346	194.9	99	280.21	0	1692
Discussant experience	346	11.6	10	8.28	0	38
Discussant co-author neighborhood	346	18.2	11	22.02	0	120
Discussant informal neighborhood	346	101.9	2	119.91	0	399

Notes: The "Presentations sample" consists of all papers with known titles presented once at Finance-related NBER SIs between 2000 and 2009. The "Journal sample" consists of all papers from the Presentation sample that were (a) published in a journal by January 2020, (b) indexed in the Scopus database and (c) discussed at most once. The "Discussants sample" consists of all papers from the Journal sample discussed once by a known discussant. *Published* equals 1 if the paper was published by January 2020. *Has discussion* equals 1 if the at least one discussant discussed the paper after its presentation. *Total citations* is the citation count as of January 2020. *Top publication* equals 1 if the paper was published in a Top 5 Economics or Top 3 Finance journal. *Journal Impact Factor* is the Scimago JIF in the year of publication. *# of pages* and *# of authors* are the count of pages and authors of the publication, respectively. *Age* is the number of years between the presentation and the publication year. *Author total Euclid* is the authors' combined Euclidean index of citations (eq. (1)). *Author total experience* is the sum of authors' experience measured as number of years since their first publication. Both variables are counted in the year before publication. The same logic applies to *Discussant* variables, which were counted in the year of the discussion. *Discussant co-author neighborhood* and *Discussant informal neighborhood* are the discussant's neighborhood centrality (eq. (2)) measured in the co-author network or the network of informal collaboration corresponding to the year of the discussion.

3 Having a Discussant and Academic Success

To study the relationship between having a dedicated discussant (versus having a general discussion with the audience only) at the NBER SIs and the academic success of paper i, we estimate the following model:

$$\text{Success}_i = \alpha_0 + \alpha_1 \cdot Paper_i + \alpha_2 Author_{i,t-1} + \beta Discussion_i + \epsilon_i, \tag{3}$$

where $Success$ is one of our main dependent variables: total citation count; whether the paper was published in a top journal; and the SCImago Journal Impact Factor of the journal where the paper was published. $Paper$ is a vector of paper-specific variables that contains the number of authors and the number of pages (of the publication).[22] $Author$ contains the joint author characteristics in $t-$: the sum of all authors' Euclidean indices of citations (See Equation (1)), the sum of individual experiences, and the square thereof. $Discussion$ is a binary variable equal to 1 if the manuscript was discussed at the workshop. In each specification we cluster standard errors on the NBER group to which the paper belongs.

Since the variable "has discussion" perfectly correlates with the NBER group, we cannot control for NBER group-fixed effects. These fixed effects would also control for field-specific effects. We deal with this lack of field-specific controls in two separate ways. One, we use JEL categories of the working paper versions or published papers in our sample. There are 540 papers for which we know the JEL category. And second, we restrict the sample to the topically most similar workshops. The journal reference cosine analysis in Table 5 indicates that the groups EFCE, EFEL, IFM and ME are most similar to each other because they have the highest pairwise overlap of weighted references to journals. We thus estimate model (3) using 279 presentations from workshops of these groups only.

Table 3 presents the β coefficients of three econometric estimations corresponding to model (3) in three specifications. The first panel uses citation counts in a negative binomial regression. While the effect of a discussion on citation count is weakly statistically significant ($p = 0.052$), the coefficient becomes statistically insignificant once controlling for topic through JEL cate-

[22] One might want to exclude $Paper$ from the regression because it is set by the editor and thus reflects quality: Better papers would be granted more space. Excluding this variable indeed drives coefficient size upwards, except in the monetary subset.

Table 3: The effect of discussions on a paper's academic success.

PANEL A - Total citations, negative binomial regression

	(1)	(2)	(3)
Has discussion	0.155	0.026	0.195
	$p = 0.052$	$p = 0.828$	$p = 0.077$
Constant	3.431	3.538	2.863
	$p = 0.000$	$p = 0.000$	$p = 0.000$
Sample	Journal	Journal	Journal, monetary workshops
Paper characteristics	✓	✓	✓
Author controls	✓	✓	✓
JEL categories		✓	
Publication year-FE	✓	✓	✓
N	592	540	279
AIC	6,677.697	6,102.018	3,035.803

PANEL B - Top journal status, logistic regression

	(1)	(2)	(3)
Has discussion	1.266	1.075	0.492
	$p = 0.000$	$p = 0.001$	$p = 0.070$
Constant	−2.688	−1.887	−2.679
	$p = 0.00000$	$p = 0.003$	$p = 0.0003$
Sample	Journal	Journal	Journal, monetary workshops
Paper characteristics	✓	✓	✓
Author controls	✓	✓	✓
JEL categories		✓	
N	592	540	279
AIC	678.509	607.747	359.969

PANEL C - Journal impact factor, OLS regression

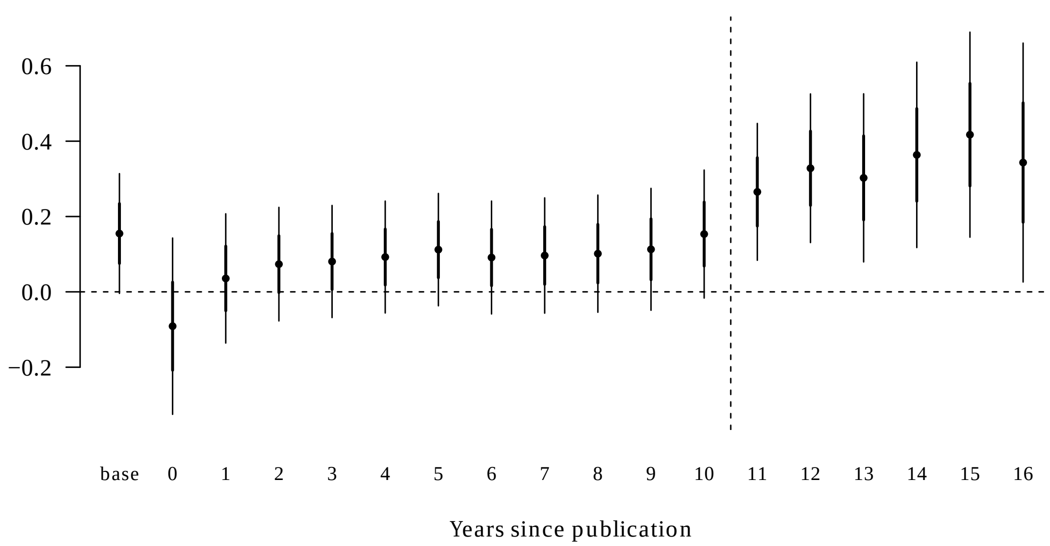
	(1)	(2)	(3)
Has discussion	2.062	2.063	1.005
	$p = 0.003$	$p = 0.025$	$p = 0.065$
Constant	−0.042	0.968	0.061
	$p = 0.971$	$p = 0.397$	$p = 0.901$
Sample	Journal	Journal	Journal, monetary workshops
Paper characteristics	✓	✓	✓
Author controls	✓	✓	✓
JEL categories		✓	
N	571	520	270
Adjusted R^2	0.257	0.277	0.181

Notes: Standard errors are clustered on NBER group. Reported coefficients in columns (1) and (2) are marginal effects. See Section 2 for variable definition. Table uses the Journal sample.

gories. Within the topically most focused subsample of monetary workshops, the coeffcient remains weakly statistically significant. The effect vanishes when we include journal fixed effects to account for the journal average of papers coming from the NBER SI workshops.

The second panel estimates a logit model to assess the relationship with top journal status. In all specifications the coefficient is statistically significant (p is at most equal to 0.07). Finally the last panel estimates an OLS model with the Journal Impact Factor as dependent variable. Similar to the previous panel we find all measures to be statistically significant. The effect is economically significant as well, as the difference in the impact factors of journals where articles with and without a discussant are being published (2.062, see Table 3) roughly equals the difference in the impact factors of *Econometrica* (19.932) and *The Journal of Finance* (18.318) in 2017.

Figure 3: The effect of having a discussant on citation counts in different years.



Notes: This figure depicts coefficients for having a discussant and citation counts at various lags. Estimation corresponds to model (3) with controls for paper and author characteristics and standard errors are clustered on the NBER group. Error bars indicate the 95% confidence intervals. Right of the dashed line less than 80% of observations remain.

As an alternative to the citations specification, we compare citation count after a fixed number of years past publication without publication year fixed effects. Figure 3 plots coefficients of "has discussion" and citation counts for different years past publication. The first dot ("base") represents the coefficient shown in model (1) of Table 3. Naturally, the more years that are included in the citation count, the fewer observations remain; right of 10 years past publication we lose more than 20% of the original sample, making a comparison difficult. Incidentally, this is also the number of years beyond which published presentations with and without discussants start to differ significantly in terms of citation counts. It seems that discussants pay off in the long run, but due to decreasing sample sizes, caution is warranted.

The analysis in this section shows that having a discussant indeed correlates with getting published in better journals and that higher citation counts of discussed papers can be explained by a journal fixed effect. One hypothesis to explain this finding is that discussants add value to the academic production and dissemination processes, but we discuss the question of causality in detail in Section 5.

Besides an improvement in quality, an alternative explanation for our observed correlation is editorial bias. In this explanation, discussants who are editors themselves treat a paper preferentially since they already know of it. To study this channel in more detail, we collect editorial tenures for discussants. Alas, no more than 14 publications were published in journals where one of the managing or associate editors discussed the same paper at the NBER SI. No valid statistical inference is possible based on this small sample size but the sheer fact that so few papers were published in a journal where the discussant is also an editor is an indication that editorial bias is an unlikely explanation.

4 Prolific Discussants and Academic Success

While the previous section focused on the extensive margin (i.e. having a discussant at an NBER SI or not), this section focuses on the intensive margin and examines different mechanisms through which discussants could impact the paper's academic success. Before going into further detail with regards to the question of causality in Section 5, we first focus on two specific types of channels through which discussants could affect the publication success of a paper: quality improvement and diffusion.

4.1 The quality-improvement channels

A discussant can improve the quality of a paper through various channels. The most immediate channel is that more prolific discussants would give better discussions that improve the paper more, while more experienced authors know the literature in which to position a paper better. We measure a discussant's prolificness using the Euclidean index of citation discussed in Section 2.3 and experience as the number of years since their first publication. We expect these effects to be reflected in all three measures of academic success of papers.

We use the Discussants sample, that is, 372 papers that were discussed exactly once by a known discussant at an NBER SI. The econometric specification relates paper success to discussant characteristics in the year of the discussion z, controlling for author characteristics in $t-1$ and paper characteristics:

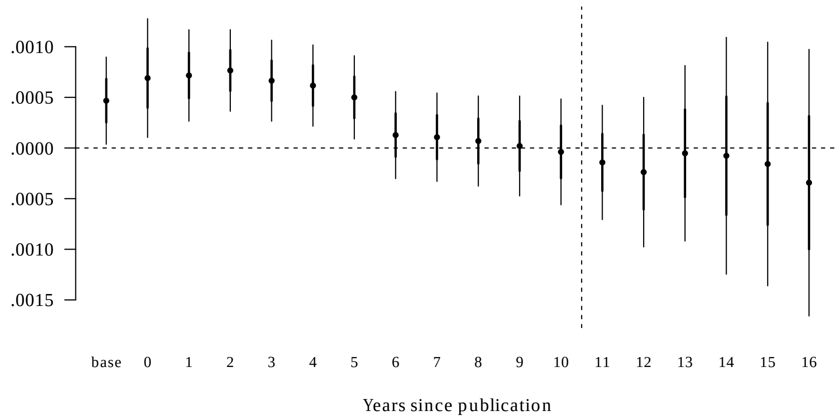
$$\text{Success}_{i,t} = \alpha_0 + \alpha_1 \cdot Paper_i + \alpha_2 Author_{i,t-1} + \beta Discussant_{i,z\leq t} + \epsilon, \tag{4}$$

where $Success$ is, as before, one of our main dependent variables: citation count; whether the paper was published in a top journal; and the Journal Impact Factor. $Paper$ is again a vector of paper-specific variables that contain the number of authors, the number of pages and the number of discussants. $Author$ contains the joint author characteristics in $t-1$: the sum of all authors' Euclidean indices of citations (See Equation (1)), the sum of individual experiences and the square thereof. $Discussant$ is a vector containing discussant characteristics as measured in the year of the discussion z: The Euclidean index of citations of the discussant, the experience of the discussant and the square thereof. In each specification we add dummies for the NBER group(s) and cluster standard errors on the NBER group to which the paper belongs.

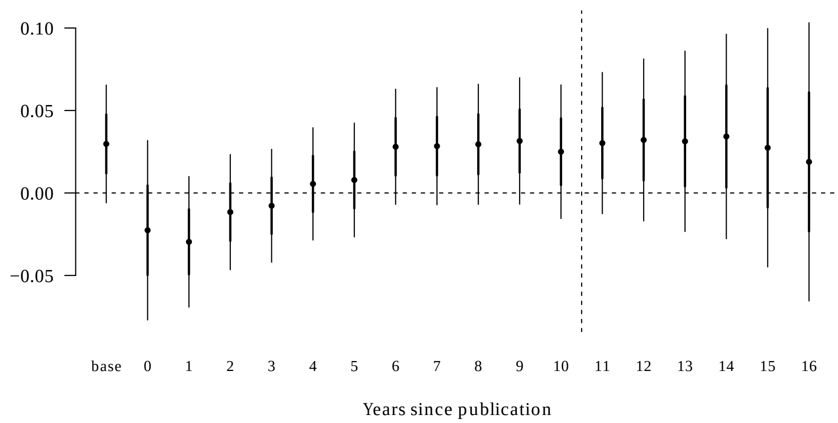
Table 4 reports results of three regressions corresponding to model (4). Column (1) estimates a negative binomial regression with total citation count as dependent variable. Column (2) estimates a logistic regression with the dependent variable whether or not the paper was published in a top journal. Both present marginal effects evaluated at the sample mean and give the percentage increase in response to a unit increase of the explanatory variable. For a one standard deviation increase in the discussants' Euclidean index of citations (280.57) from the mean, the expected citation count for the average paper increases by $280.57 \times 0.0005 \times 100 \approx$

Figure 4: The effect of discussant characteristics on citation counts in different years.

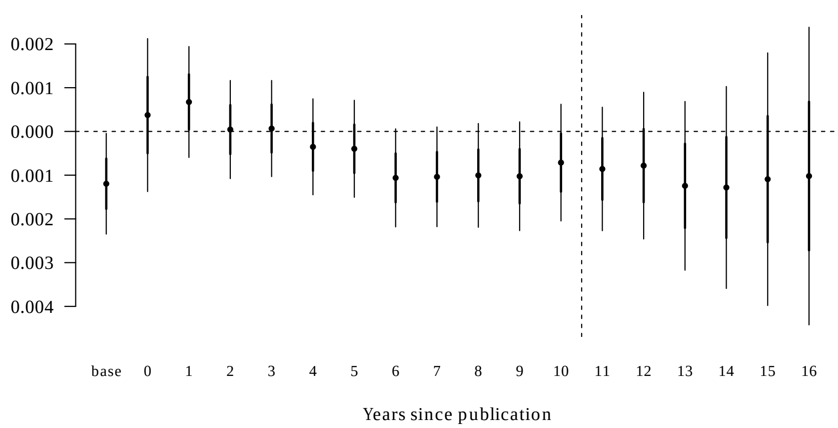
(a) Discussant Euclidean index



(b) Discussant experience



(c) Discussant experience^2



Notes: This figure uses the Discussant sample. Plots depict coefficients for various discussant characteristics and citation counts at various years past publication. Estimation corresponds to model (4) with controls for paper and author characteristics, with NBER group dummies and standard errors clustered on the NBER group. See section 2 for variable definitions. Error bars indicate the 95% confidence intervals. Right of the dashed line less than 80% of observations remain.

Table 4: The effect of discussant characteristics on a paper's academic success.

	Total citations	Top publication	Journal Impact Factor
	neg. bin.	*logistic*	*OLS*
	(1)	(2)	(3)
Discussant Euclid	0.0005	0.001	0.003
	$p = 0.030$	$p = 0.337$	$p = 0.000$
Discussant experience	0.030	0.038	−0.061
	$p = 0.096$	$p = 0.504$	$p = 0.331$
Discussant experience^2	−0.001	−0.001	0.001
	$p = 0.038$	$p = 0.495$	$p = 0.641$
Constant	4.064	−2.506	2.126
	$p = 0.000$	$p = 0.040$	$p = 0.297$
Paper characteristics	✓	✓	✓
Author characteristics	✓	✓	✓
NBER group-FE	✓	✓	✓
Publication year-FE	✓		
N	346	346	339
Adjusted R^2			0.328
AIC	3,923.387	332.032	

Notes: Standard errors are clustered around NBER working group. Reported coefficients in columns (1) and (2) are marginal effects. See Section 2 for variable definitions. This table uses the Discussants sample.

14.03% from the mean, which translates into $0.1403 \times 138.1 \approx 19$ citations over the paper's lifetime. There is, on the other hand, only a weakly statistically significant relationship with a discussant's experience and citation count. Column (3) reports OLS coefficients with the Journal Impact Factor as dependent variable. We find statistically significant coefficients of discussants' productivity with the paper's citation count and the journal quality as measured by its impact factor.

Figure 4 relaxes the citations model in column (1) and presents coefficients of different specifications with citation counts only up to specific years and without publication year fixed effects. Panel 4a shows that the discussant's prolificness (holding other characteristics constant) impacts short-term citation counts (up to 5 years past publication) the most and intermediate and late citation counts not at all. With discussants' experience (Panel 4b) we find the opposite relationship. The coefficient for a discussant's experience is negative in the short-term citation counts (although imprecisely measured) and only becomes positive beyond 6 years. This comes at a decreasing rate (Panel 4c).

4.2 Diffusion channel

While there is evidence consistent with a quality-improving channel, discussants might improve a paper's citation count due to diffusion of knowledge about the paper. To test for such a diffusion channel, we examine the relationship between a discussant's ability to diffuse the paper and our measures of publication success. It is reasonable to expect this channel to directly affect the citation count, given evidence on the diffusion of technology, for example, in farming in emerging markets (Conley and Udry, 2010) or in microfinance (Banerjee et al., 2013). Specifically, by being the first to learn about the existence and quality of the paper, discussants may not only decide to cite the paper themselves, but also let their immediate collaborators know about it. We understand collaborators in a broader sense, including both co-authors and informal collaborators. This diffusion can happen through regular discussions or through an indirect process where a discussant's collaborators learn about the paper because the discussant cites it. In light of this, we examine two hypotheses:

(i) Discussants are among the first to cite a paper and they cite a significant number of the papers they discuss.

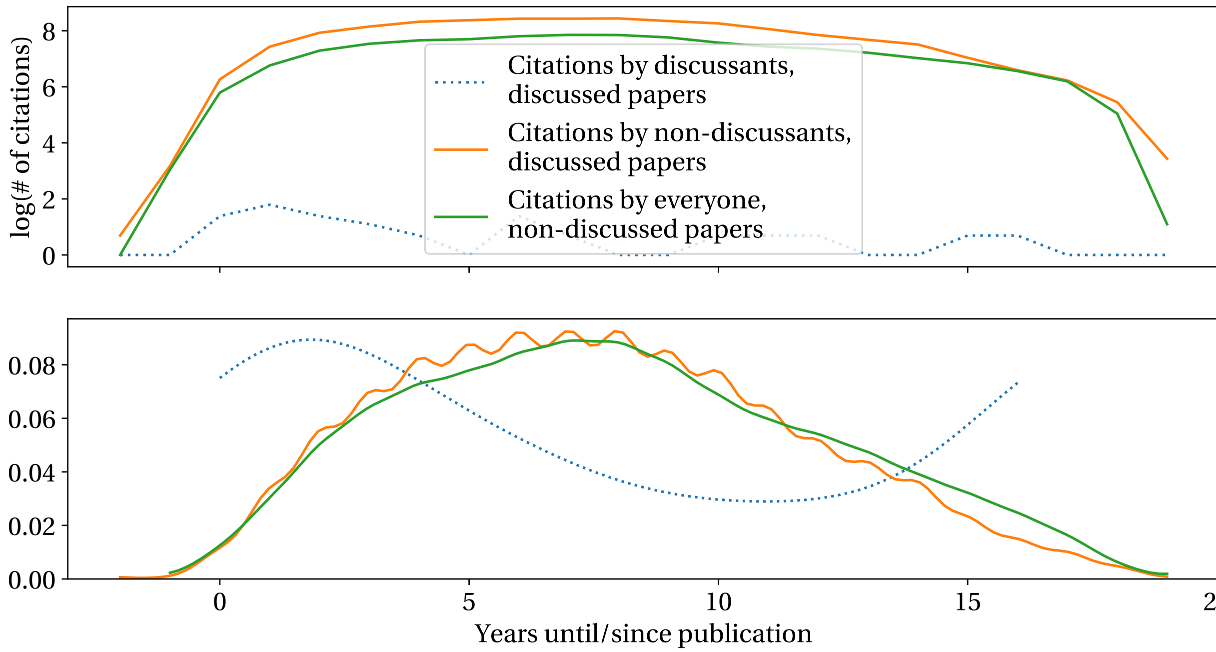
(ii) A discussant's ability to diffuse the paper to the scientific community positively correlates with the paper's citation count.

To test the first hypothesis, we compare the citation patterns of discussants and non-discussants. Figure 5 compares the annual citation stream from discussants and non-discussants to both discussed papers and non-discussed papers.[23] The upper panel shows the log number of citations, and the lower panel shows the corresponding kernel density estimate. The blue dotted lines are based on 23 citations by 15 discussants of 19 papers (out of 345). Those 345 papers

[23]We restrict ourselves to citations in publications recorded in Scopus.

received in total 47,636 citations by January 2020. While it becomes clear that the majority of discussants do not cite the paper they discuss, we cannot say anything about the unconditional probability that a discussant would have cited the paper without having discussed it. This counterfactual is simply unobservable.[24]

Figure 5: Citation streams to discussed and non-discussed papers, by status of discussant.



Notes: These figures show citation patterns for discussed and non-discussed papers in the Journal sample by status of citing author (discussant or not). The upper panel gives the log number of citations by year, as recorded from published papers recorded in Scopus; the lower panel gives the corresponding kernel density estimations.

While the small number of incidences indicates that the direct diffusion effect is probably negligible, another fact is at odds with the first hypothesis. If discussants cite a paper, they do so relatively soon after publication. Yet, non-discussants cite the paper in early stages as well—just not the majority of citing non-discussants. Discussants also stop citing "their" paper earlier.

To test the second hypothesis, we use neighborhood centrality as defined in Section 2.5. We compute the centrality separately in the co-author network and the network of informal intellectual collaboration corresponding to the year of the discussion z. The relevant networks thus

[24]Given our survey of NBER SI workshop organizers this is not entirely surprising: Most editors indicated that they are looking for discussants who provide a new perspective on a paper (i.e. they are not necessarily looking for discussants who work on exactly the same topics as the author).

use papers published in $z, z+1, and z+2$. The econometric model relates a paper's total citation count to the discussants' neighborhood centrality controlling for discussant characteristics in year z, for author characteristics in $t-1$ and paper characteristics. There are dummies for the corresponding NBER groups, on which we also cluster standard errors:

$$\begin{aligned} \text{Total Citations}_{i,t} &= \alpha_0 + \alpha_1 \cdot Paper_i + \alpha_2 Author_{i,t-1} + \alpha_3 \cdot Discussant_{i,z<t} \\ &\quad + \beta \cdot Neighborhood_{i,z<t} + \epsilon_{i,t}. \end{aligned} \tag{5}$$

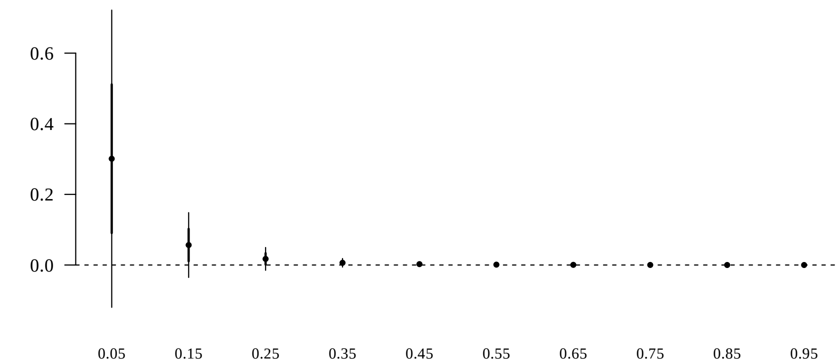
In case β does not differ from zero, we reject hypothesis (ii) stating that increased citation counts are due to discussants' diffusion efforts.

Figure 6 plots the coefficients of neighborhood centrality using various values of $\delta \in [0.05, 0.15, \ldots, 0.85, 0.95]$. The top panel uses neighborhood centrality computed in the co-author network, the bottom panel uses neighborhood centrality computed in the network of informal collaboration. None of the coefficients differs significantly from zero. We therefore reject the hypothesis that discussants with a higher ability to diffuse a paper (i.e. neighborhood centrality) increase citation count.

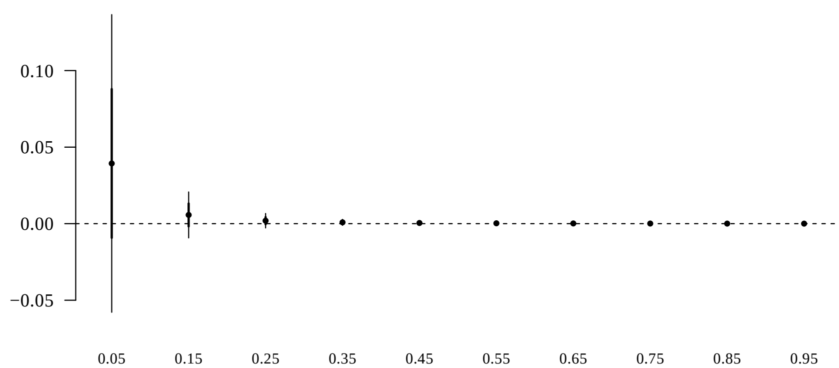
A plausible explanation for the lack of evidence of a diffusion channel is that alternative methods through which researchers find papers (e.g. internet search or simply going to the library) are just as important as learning about new papers through social networks.

Figure 6: Neighborhood centrality and total citation count.

(a) Co-author networks



(b) Informal networks



Notes: These figures depict coefficients for neighborhood centrality of discussants with different values of $\delta \in [0.05, 0.15, \ldots, 0.85, 0.95]$ for two different networks. Estimation corresponds to model (5) with controls for paper, author and discussant characteristics, with NBER group dummies and with standard errors clustered on the NBER group. The top panel shows coefficients for a discussant's neighborhood centrality computed in the co-author networks. The bottom panel shows coefficients for discussant's neighborhood centrality computed in the networks of informal collaboration. Error bars indicate the 95% confidence intervals.

5 Causality and Challenges to Identification

So far, we have identified a correlation between whether a paper has a discussant and the discussant's characteristics on the one hand, and the paper's publication success on the other. But our results do not establish causality per se (i.e. through randomized assignment of discussants), and our setting poses a number of challenges to causality. In an ideal setting, papers would be allocated randomly first to workshops with or without discussants, and in the former case randomly to discussants. This is of course not the case. In this section we discuss assumptions under which our results can be interpreted as causal evidence, regardless.

5.1 Workshops with and without discussants

Whether a paper is discussed at a workshop during an NBER SI depends on two consecutive choices: The authors apply to a workshop that features discussants, and the corresponding organizers accept the paper. Thus, neither of the two decisions makers engages in sorting. The kind of sorting that prevents identification is one where either submission or acceptance is based on the fact that there are (or are not) discussants.

Our identifying conditions are thus: (i) neither authors nor organizers sort into workshops based on the fact that they have discussants; (ii) papers are of comparable quality (e.g. organizers do not accept papers of different quality); and (iii) the workshops differ in nothing but the existence of discussants.

We argue that the first condition holds because it is rational for authors to seek a good topical fit with the workshop's theme. This is corroborated by the fact that some papers were presented both in workshops with and without discussants, which requires that these authors applied to both types of workshops (when applying for the SI workshops authors may state up to three workshops that they wish to present at). This is also supported by the findings of the short survey we conducted. If there are strategic considerations beyond seeking a good topical fit, it is more reasonable for authors to apply for workshops that editors of their target journals attend.

To show that manuscripts at the NBER workshops are of comparable quality, we rely on two noisy proxies: the average affiliation rank of the authors' institutions and the readability of the abstract of the manuscript. Both measures are computed at the year of the workshop. Manuscript readability, in particular, has been shown to be an indicator of future citation counts (Dowling et al., 2018).

For the affiliation rank, we use the Tilburg University Economics Ranking of the authors' first affiliations as stated in the program.[25] The upper right panel of Figure 7 shows that there are no statistically significant differences between papers with and without discussants: Both sets of papers are written by equally ranked author groups. This holds whether one compares the individual manuscripts (left part) or the average manuscripts by workshop (right part).

The lower two plots of Figure 7 show that abstracts at the time of the workshop score equally on readability indices. Here we use two measures of readability, the Gunning Fog Index and the Simple Measure of Gobbledygook (SMOG). Hence pre-discussion quality, as measured by readability, is the same for manuscripts that get discussed and those that do not. Both scores are the number of years necessary to understand the text. The Gunning Fog Index takes into account the number of words per sentence and the share of complex words over the total word count.[26] The SMOG focuses solely on lengthy words (those with three or more syllables, called polysyllables) and their relation to the number of sentences.[27]

Another confounding factor, to which the third assumption speaks to, arises if the workshops with discussants differ from workshops without discussants in other dimensions. We probe two dimensions: duration of presentation and topical overlap. The upper left plot of Figure 7 shows that at least for the duration of the presentation, both groups are on average the same, namely 55 minutes. This holds whether one compares the duration of individual presentations (left part of the panel) or the average duration of presentations by workshop (right part of the panel).

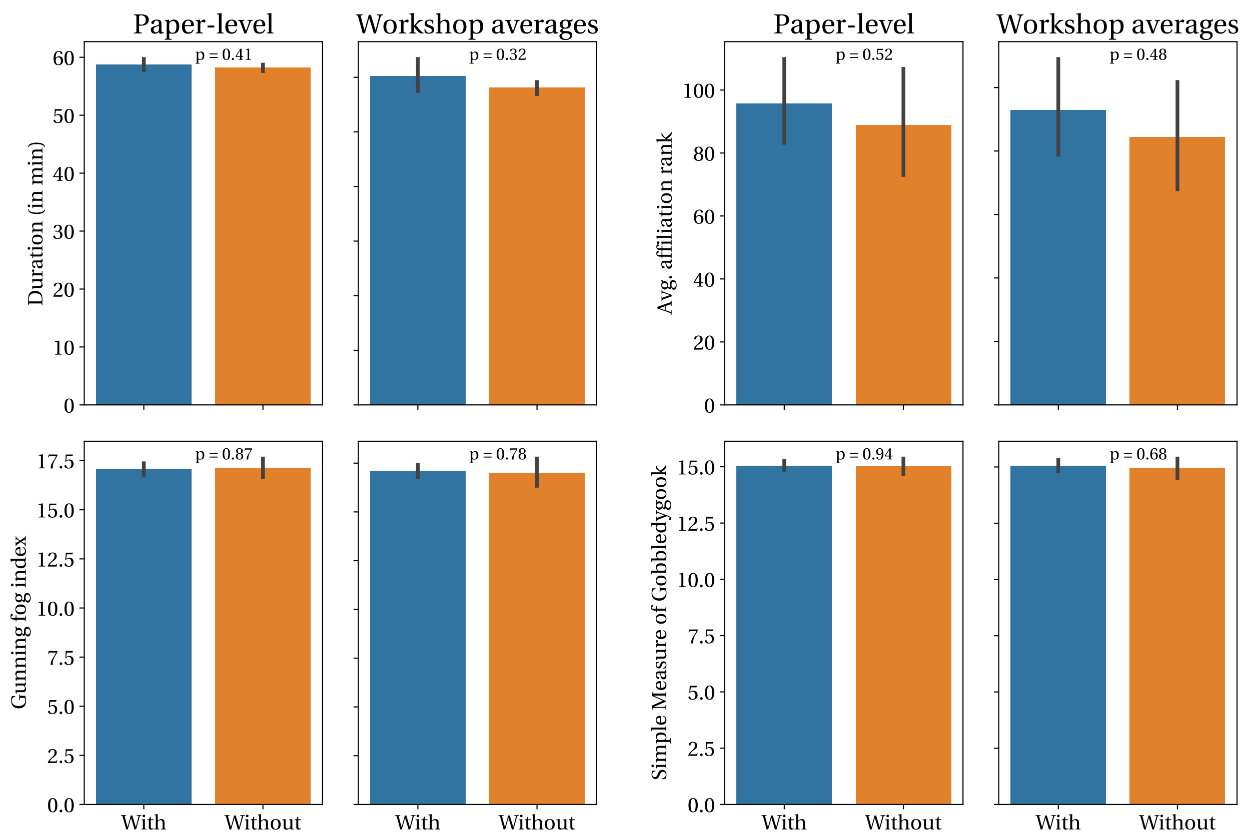
The hosting NBER groups and programs are furthermore topically similar. Table 5 presents the pairwise similarities of group and programs in our sample obtained from cited journals of the published papers. Similarity is defined as one minus the cosine of two vectors. Each vector represents all aggregated reference lists of all published manuscripts presented at workshops of that group. Each vector's element is the weighted count of a particular journal, where we ob-

[25]For details and the rankings see https://econtop.uvt.nl/methodology.php.

[26]The formula is $0.4\left[\frac{\text{word count}}{\text{sentence count}}+100\frac{\text{complex word count}}{\text{word count}}\right]$. Computed using *textatistic*.

[27]The formula is $3.1291+1.0430\times\sqrt{30\times\frac{polysyllablescount}{sentencecount}}$. Computed using *textatistic*.

Figure 7: The comparability of manuscripts with and without discussants.



Notes: These barplots show the differences between papers with and without discussants according to different dimensions. The Paper level panel uses values by paper, while the Workshop average panel averages paper values by workshop first and then compares those means by group category. The annotated p-values correspond to the t-statistics testing equal means. Dimensions are (in clockwise order) the following: the length of the allocated time for a presentation (including discussion), the average Tilburg University Economics Ranking of the authors of the presented manuscript in the year of the discussion, the readability according to SMOG of the abstract of the presented manuscript, and the readability according to the Gunning Fog Index of the abstract of the presented manuscript.

tained weights from tfidf-vectorization on the entire groups×journals matrix.[28] This vectorization gives higher weights to rare occurrences and is a standard practice in natural language processing. The similarities are bound between 0 and 1, where higher values indicate higher similarity. The average similarity across all groups is 0.54. If we pool all groups with and without discussants, such that we obtain only two vectors, the similarity increases to 0.62. This value

[28]The formula for each entry g, j in the groups×journals matrix with n journals obtains as $f_j \times \mathrm{idf}(t)$, where f_t is the frequency count of journal j $\mathrm{idf}(j) = \log \frac{1+n}{1+|\{g \in D : j \in g\}|} + 1$.

Table 5: Similarity matrix of NBER groups and programs.

	AMRE	**AP**	**CF**	EFCE	**EFEL**	EFWW	**IFM**	ME	**PERE**	**RISK**
AMRE	1									
AP	0.57	1								
CF	0.55	0.95	1							
EFCE	0.33	0.32	0.29	1						
EFEL	0.58	0.83	0.8	0.7	1					
EFWW	0.32	0.44	0.36	0.4	0.48	1				
IFM	0.37	0.42	0.43	0.8	0.71	0.42	1			
ME	0.35	0.32	0.32	0.95	0.68	0.44	0.8	1		
PERE	0.68	0.25	0.28	0.41	0.45	0.21	0.39	0.42	1	
RISK	0.56	0.88	0.86	0.47	0.87	0.41	0.54	0.48	0.32	1

Notes: This table shows pairwise similarities between the NBER groups and programs used in our sample. The NBER groups and programs are "Asset Marketing/Real Estate" (AMRE), "Asset Pricing" (AP), "Corporate Finance" (CF), "Impulse and Propagation Mechanisms" (EFCE), "Capital Markets in the Economy" (EFEL), "Forecasting and Empirical Methods in Macro and Finance" (EFFE), "International Finance and Macroeconomics" (IFM), "Monetary Economics" (ME), "Finance and Macro" (MEFM), "Economics of Real Estate and Local Public Finance" (PERE) and "Risk of Financial Institutions" (RISK). Groups or programs printed in bold include discussants. Similarities are 1 minus the cosine similarity of vectors representing the NBER groups. Each vector is inferred from the cited journals of the published manuscripts presented in the corresponding workshop. The entries in the vector correspond to weighted counts of how often a given journal was cited, where the weights are obtained from tfidf-vectorization on the entire groups×journal matrix.

is close to the similarity of only groups with discussants (0.63), or to the similarity of only groups without discussants (0.61).

A final concern arises from different behaviour of authors before the workshop depending on whether they get or do not get a discussant. For example, authors may polish the paper more when they know a discussant is going to carefully read it. However, the same is true for workshops without discussants: 73% of manuscripts presented in workshops without discussants have a version online and linked in the workshop program. The corresponding figure for manuscripts presented in workshops with discussants is 84%. That the two figures are so close indicates that authors generally strive to have a version of their paper online by the time of the workshop, irrespective of whether or not they will have a discussant. Since readability of abstracts is the same for manuscripts in workshops with and without discussants, it is also unlikely that authors put more effort into its preparation just because there is a discussant. This would speak to a deadline effect (Bonatti and Hörner, 2011).

5.2 Discussant characteristics

Which discussant a paper ultimately receives depends on the workshop organizers. Crucially for our analysis, authors are not involved in the selection process. Since the assignment is non-random, we would need additional assumptions to identify the effect of discussant characteristics: (i) organizers match discussants to papers and authors based primarily on topical fit (e.g. whether the discussant knows the relevant literature, methods and data) rather than discussant characteristics (e.g. prolificness, age and gender); and (ii) discussants discuss a paper irrespective of the session it is in (i.e. no sorting of discussants). Formally, the covariance between discussant characteristics on the one side and author as well as paper characteristics on the other side must be zero.

The survey we conducted among workshop organizers sheds further light on the matching of discussants to papers. Organizers have confirmed that discussants rarely reject an invitation. Being a discussant at an NBER SI is considered a prestigious opportunity, thus, invited discussants rarely decline. Consequently, the second condition holds. Responding organizers unanimously said they try to look for discussants that can ignite a lively debate. Indeed, the ability to present well, summarize the paper and relate the results to the existing literature are more important than experience and prolificness, for example. Specifically, no organizer indicated any form of assortative matching between prolific discussants and promising papers, which is arguably the biggest threat to our identification strategy. In fact, some organizers explicitly indicate that they are looking for discussants who are not too close to the authors and bring a new perspective.

To quantitatively study the existence of assortative matching in our sample, we look for statistically significant relationships between the observable discussant characteristics and observable author and paper characteristics. The discussant characteristics we check are the Euclidean index of citations, experience, neighborhood centrality in the co-author network, and neighborhood centrality in the network of informal collaboration. Each variable is measured in the year of the discussion (or in the network corresponding to the year of the discussion).

Explanatory variables are the authors' joint Euclidean index of citations, their joint experience, the square thereof and the number of authors.[29]

Table 6: Assortative matching of discussant characteristics to author and paper characteristics.

	Discussant experience	Discussant Euclid	Discussant co-author neighborhood	Discussant informal neighborhood
	Neg. Bin.		*OLS*	
	(1)	(2)	(3)	(4)
Author total Euclid	−0.00002	−0.026	−0.001	0.002
	$p = 0.863$	$p = 0.369$	$p = 0.836$	$p = 0.936$
Author total experience	0.006	−2.059	0.389	−2.693
	$p = 0.440$	$p = 0.380$	$p = 0.432$	$p = 0.114$
Author total experience^2	−0.00002	0.079	−0.003	0.046
	$p = 0.870$	$p = 0.039$	$p = 0.752$	$p = 0.098$
# of authors	−0.043	3.054	2.940	21.565
	$p = 0.450$	$p = 0.863$	$p = 0.429$	$p = 0.093$
Constant	2.563	218.128	8.761	148.178
	$p = 0.000$	$p = 0.034$	$p = 0.686$	$p = 0.048$
Discussion year-FE	✓	✓	✓	✓
NBER group-FE	✓	✓	✓	✓
N	441	441	441	441
Adjusted R^2		0.152	0.198	0.335
AIC	3,064.316			

Notes: See Section 2 for variable definitions. Samples include both published and unpublished manuscripts.

Table 6 reports the results of this exercise using the Presentations sample. That is, we include all presentations, even those that did not result in a publication.[30] There are but two significant relationships that we find. One is between authors' squared experience and a discussant's Euclidean index of citations. Since the relationship with the authors' total experience

[29] We assume the author group size to be constant between the discussion and the publication. We exclude the number of pages here because we do not observe the number of pages prior to publication such that it would be comparable across papers. However, including the number of pages of the published paper does not change the results.

[30] When restricting our sample to manuscripts that resulted in a publication, coefficients change very little and become *less* significant.

is insignificant, this matching does not raise concerns. We also find that papers with larger author groups tend to receive discussants that are more neighborhood central in the network of informal collaboration. Since this variable has no statistically significant effect on the paper's citation count, it does not pose challenges to our identification either. In any case, the finding here only concerns the second set of results, but not the first (where we focus on the existence of a discussant). In total, Table 6 alleviates the fear of assortative matching.

Finally, a confounding factor arises from the differing ability of the organizers to attract discussants. Some organizers might, due to their standing or prominence in the field, reach a wider pool of researchers that could discuss a paper. But organizers tend to be highly regarded and well-connected academics who know their respective field well beyond their personal network. Varying reach within the profession might thus not be a severe issue. Since organizers seldom change during our sample period, the NBER group fixed effects we introduce however take the organizers' time-invariant ability to attract discussants into account. Additionally, we cluster standard errors on the NBER working group (or joint session) to account for unobserved heterogeneity, for example, in the form of the organizer's professional network.

5.3 Publication bias

Finally, we face a selection bias coming from the fact that we only observe published papers in the Journal sample and in the Discussants sample. It might be that discussed papers have a lower propensity to be published and those that get published, get published in highly prestigious journals. Here the interpretation of the discussant's role is that of a gatekeeper or screening device, which would bias our results. However, the opposite seems to be true: Discussed papers have a higher, not lower, propensity to be published. Table 7 uses the 845 observations from the Presentations sample, of which 630 (75%) had a discussion. The relationship doesn't change if we control for pre-publication as a NBER working paper, although the model's goodness-of-fit increases drastically.

Table 7: The effect of discussions on publication status.

	Published	
	(1)	(2)
Has discussion	0.562	0.384
	$p = 0.0005$	$p = 0.024$
NBER Working Paper		17.853
		$p = 0.968$
Constant	0.791	0.458
	$p = 0.000$	$p = 0.0002$
N	846	846
AIC	944.158	798.183

Notes: NBER Working Paper equals 1 if the manuscript is available as a NBER Working Paper. See Section 2 for definitions of other variables. Table uses the Presentations sample.

6 Conclusion

A handful of papers study informal collaboration and find positive correlation between the extent of informal collaboration (i.e. the number of acknowledged commenters, seminars and conferences) and citation count. Insightful as these seminal findings are, they are not specific for discussants, nor do they tell us much about some underlying mechanisms.

Our setting—discussions at Finance-related NBER Summer Institutes—comes much closer to causal inference and allows us to study the underlying mechanisms. Papers benefit from having a discussant in terms of an increased probability of being published in prestigious journals. Because discussants read and discuss the paper in detail their suggestions for improvements will likely differ from the audience's suggestions. Papers' citation counts increase in the prolificness of discussants, but not in their ability to diffuse the paper. This finding lends support to the existence of a quality-improving channel of informal collaboration.

Our findings are relevant for the design of conferences that are designed to not only showcase the latest research, but also to provide authors with useful feedback from the audience and, often, discussants. Since having a discussant correlates with a higher probability of subsequent publication in a top journal, conference organizers should consider having a dedicated discussant for each paper. Yet for a valid economic analysis of this question we lack the knowledge of the cost of informal collaboration.

Lastly, our insights about the role of dedicated discussants contributes to the literature studying processes of knowledge production. Understanding them is highly relevant as economies shift towards the production of intangible and increasingly knowledge intensive goods.

References

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

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

Bandiera, O. and Rasul, I. (2006), 'Social Networks and Technology Adoption in Northern Mozambique', *The Economic Journal* **116**(514), 869–902.

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

Baruffaldi, S. and Pöge, F. (2020), 'A Firm Scientic Community: Industry Participation and Knowledge Diffusion', *Max Planck Institute for Innovation & Competition Research Paper* (20-10).

Belenzon, S. and Schankerman, M. (2013), 'Spreading the Word: Geography, Policy, and Knowledge Spillovers', *Review of Economics and Statistics* **95**(3), 884–903.

Bonatti, A. and Hörner, J. (2011), 'Collaborating', *The American Economic Review* **101**(2), 632–663.

Borjas, G. J. and Doran, K. B. (2015), 'Which Peers Matter? The Relative Impacts of Collaborators, Colleagues, and Competitors', *Review of Economics and Statistics* **97**(5), 11041117.

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

Castaldi, S., Giacometti, M., Toigo, W., Bert, F. and Siliquini, R. (2015), 'Analysis of full-text publication and publishing predictors of abstracts presented at an Italian public health meeting (20052007)', *BMC Research Notes* **8**(1), 492.

Combes, P.-P. and Linnemer, L. (2010), 'Inferring missing citations: a quantitative multi-criteria ranking of all jornals in economics', *HAL Working Paper* (28).

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

Coupé, T., Ginsburgh, V. and Noury, A. (2010), 'Are leading papers of better quality? Evidence from a natural experiment', *Oxford Economic Papers* **62**(1), 1–11.

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.

Dowling, M., Hammami, H. and Zreik, O. (2018), 'Easy to read, easy to cite?', *Economics Letters* **173**, 100–103.

Feenberg, D., Ganguli, I., Gaulé, P. and Gruber, J. (2017), 'Its Good to Be First: Order Bias in Reading and Citing NBER Working Papers', *Review of Economics and Statistics* **99**(1), 32–39.

Gans, J. S. and Shepherd, G. B. (1994), 'How Are the Mighty Fallen: Rejected Classic Articles by Leading Economists', *Journal of Economic Perspectives* **8**(1), 165–179.

Genicot, G. and Bramoullé, Y. (2018), 'Diffusion Centrality: Foundations and Extensions', *AMSE Working Papers* (37).

Golub, B. and Jackson, M. O. (2010), 'Naïve Learning in Social Networks and the Wisdom of Crowds', *American Economic Journal: Microeconomics* **2**(1), 112–149.

Gorodnichenko, Y., Pham, T. and Talavera, O. (2019), 'Conference Presentations and Academic Publishing', *NBER Working Paper Series* (26240).

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

Heckman, J. J. and Moktan, S. (2020), 'Publishing and Promotion in Economics: The Tyranny of the Top Five', *Journal of Economic Literature* **58**(2), 419–470.

Ioannidis, J. P. A. (2012), 'Are Medical Conferences Useful? And for Whom?', *The Journal of the American Medical Association* **307**(12), 1257.

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

Merton, R. K. (1968), 'The Matthew Effect in Science: The reward and communication systems of science are considered', *Science* **159**(3810), 56–63.

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.

Oswald, A. J. (2010), 'A suggested method for the measurement of world-leading research (illustrated with data on economics)', *Scientometrics* **84**(1), 99–113.

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. and Georg, C.-P. (2018), 'What 5,000 Acknowledgements Tell Us About Informal Collaboration in Financial Economics', *Max Planck Institute for Innovation & Competition Discussion Paper 11*.

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

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

A Further tables

Table A1: Spearman and Pearson correlation coefficients for all variables in the Journal sample.

Total citations		0.36	0.27	0.28	0.13	−0.39	0.13	0.01	0.05
Top publication	0.28		0.81	0.40	0.11	−0.16	0.02	−0.08	0.28
Journal Impact Factor	0.17	0.75		0.40	0.14	−0.05	0.14	−0.01	0.24
# of pages	0.25	0.37	0.43		0.09	−0.05	0.11	−0.02	0.14
# of authors	0.09	0.09	0.13	0.08		0.05	0.50	0.58	0.13
Age	−0.26	−0.17	−0.08	−0.06	0.07		0.11	0.14	0.15
Author total Euclid	0.10	0.03	0.13	0.03	0.30	0.10		0.84	0.01
Author total experience	0.02	−0.07	0.01	−0.01	0.58	0.18	0.57		−0.02
Has discussion	0.02	0.28	0.24	0.12	0.12	0.16	−0.01	−0.01	

Notes: Upper triangular shows Spearman (rank) correlation coefficients, while lower triangular shows Pearson correlation coefficients. See Section 2 for variable definition. Table uses the journal sample.

Table A2: Spearman and Pearson correlation coefficients for all variables in the Discussants sample.

Total citations		0.39	0.28	0.24	0.06	−0.48	0.05	−0.05	−0.01	−0.07	0.05	0.05
Top publication	0.28		0.79	0.43	0.08	−0.29	0.07	−0.07	0.13	0.00	0.18	0.23
Journal Impact Factor	0.18	0.75		0.46	0.12	−0.21	0.18	−0.01	0.12	0.01	0.19	0.21
# of pages	0.21	0.42	0.48		0.13	−0.05	0.18	0.03	0.04	−0.05	0.15	0.12
# of authors	0.02	0.07	0.11	0.14		0.02	0.50	0.60	0.08	0.03	0.13	0.17
Age	−0.31	−0.30	−0.21	−0.07	0.04		0.10	0.14	0.02	0.06	−0.01	−0.04
Author total Euclid	0.03	0.09	0.17	0.06	0.23	0.11		0.82	0.21	0.14	0.14	0.11
Author total experience	−0.03	−0.06	0.02	0.05	0.57	0.20	0.54		0.13	0.12	0.10	0.05
Discussant Euclid	0.00	0.16	0.18	0.04	0.08	0.05	0.18	0.15		0.81	0.19	0.21
Discussant experience	−0.06	−0.03	−0.01	−0.07	0.01	0.13	0.17	0.12	0.54		0.04	−0.02
Discussant co-author neighborhood	0.01	0.15	0.22	0.21	0.16	−0.05	0.13	0.18	0.23	0.09		0.38
Discussant informal neighborhood	−0.02	0.22	0.21	0.10	0.16	−0.05	0.12	0.06	0.31	−0.02	0.39	

Notes: Upper triangular shows Spearman (rank) correlation coefficients, while lower triangular shows Pearson correlation coefficients. See Section 2 for variable definition. Table uses the discussants sample.

B Algorithmic steps to compute neighborhood centrality

Neighbourhood centrality counts and sums up the number of τ-order neighbours while discounting for information decay in distant neighborhoods. When applied to the diffusion of information about a newly produced paper, a scientist with a higher neighborhood centrality has a potential to diffuse information about a paper to a larger portion of the network than one with a lower neighborhood centrality. We use the following matrix-based algorithmic steps for computing neighbourhood centrality. A matrix-based algorithm has a lower time complexity than computing centralities one node at a time.

Let G be an adjacency matrix of a directed unweighted network. That is, G is a zero-one matrix with elements $g_{ij} = 1$ if there is a link from i to j and zero otherwise. Let I denote the identity matrix. For any non-negative arbitrary matrix M, let $M^{[1]}$ be a matrix derived from M by replacing all elements of M that are greater than one with one. Similarly, if M is an arbitrary matrix containing some elements that are negative, let $M^{[0]}$ be a matrix derived from M by replacing all negative elements of M with zero. These two matrix operations also extend to sums and products of matrices. For example, we write $[A \times M]^{[1]}$ for a matrix derived from the product $A \times M$ by replacing all elements of $A \times M$ that are greater than one with one. Similarly, $[A+M]^{[1]}$ is a matrix derived from $A+M$ by replacing all elements of $A+M$ that are greater than one with one.

Let $W(\tau)$, for $\tau = 1, 2, \cdots$, be the *τ-order adjacency matrix*. That is, for each τ, the ijth element of $W(\tau)$ is 1 if j is at distance τ (and not at distances $1, 2, \cdots, \tau - 1$) from i, and zero otherwise. These $W(\tau)$ matrices are computed through the following iterative steps.

First, $W(1) = G$. Second, to compute $W(2)$, first notice that the elements of the product $G \times G$ are the number of two-step paths leading up to a given node (i.e. the ijth element of $G \times G$ is the number of paths that start from i and end at j in two steps, including paths leading back to i). Thus, if all elements of $G \times G$ that are greater than one are replaced by 1, then the ijth element of $[G \times G]^{[1]}$ is one if j is reachable in two steps from i, and zero otherwise. Note that if j is a direct (i.e. first-order) neighbour of i but is reachable in two steps (i.e. there exist a path from i to some k and then to j), then the ijth element of $[G \times G]^{[1]}$ will also be one. However, by definition of τ-order neighbours, no single node should appear in more than one τ-order neighbourhood of a given node. To account for cyclic paths, we subtract $I+G$ from $[G\times G]^{[1]}$ and then replace all negative elements by zero, so that the ijth element of $[[G\times G]^{[1]}-I-G]^{[0]}$ is one if j is reachable in two steps from i but not in one step. Thus, $W(2)=[[W(1)\times G]^{[1]}-I-W(1)]^{[0]}$, where we apply the equality $W(1)=G$.

Third, following the same steps above, the ijth element of $W(2)\times G$ is the number of paths that start from i and end at j in three steps. Replacing all elements of $W(2)\times G$ that are greater than one by one yields $[W(2)\times G]^{[1]}$, whose ijth element is one if j is reachable in three steps from i, and zero otherwise. Subtracting $I+W(1)+W(2)$ from $[W(2)\times G]^{[1]}$ and replacing all negative elements by zero yields $W(3)=[[W(2)\times G]^{[1]}-I-W(1)-W(2)]^{[0]}$. Applying the same steps iteratively, we find that after t iterations,

$$W(t)=\left[[W(t-1)\times G]^{[1]}-\sum_{\tau=0}^{t-1}W(\tau)\right]^{[0]}$$

where $W(0)=I$.

Now, let $\mathbf{e}$ be a column vector of ones, and let T be the value of τ at which all elements of $W(\tau)$ become zero. Then the vector of neighborhood centralities with discount factor δ is calculated as

$$n(\delta)=\sum_{\tau=1}^{T}\delta^{\tau}W(\tau)\mathbf{e} \tag{6}$$

Consider an example of a network in Figure A1 with its respective adjacency matrix G. After $T=9$ steps of iteration, the $W(\tau)$ matrix converges to zero. The number of iterations amounts to $d(G)+1$, where $d(G)$ is the diameter of the network (for the network in Figure A1, the diameter corresponds to the distance from node 1 to node 13). Using eq. (6), we can compute the neighbourhood centralities vector for various values of discount factor δ. For example, when $\delta=0.8$, our computations yield:

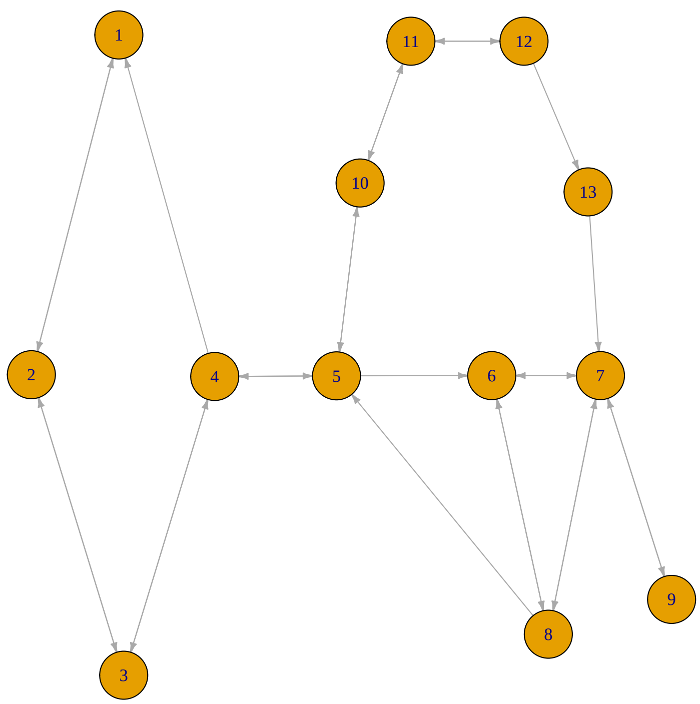
$$
\begin{aligned}
\delta(0.8) &= \sum_{\tau=1}^{9} 0.8^{\tau} W(\tau)\mathbf{e} \\
&= 0.8\begin{pmatrix} 1\\2\\2\\3\\3\\2\\3\\3\\1\\2\\2\\2\\1 \end{pmatrix} + 0.8^{2}\begin{pmatrix} 1\\1\\2\\3\\5\\2\\1\\3\\2\\3\\2\\2\\3 \end{pmatrix} + 0.8^{3}\begin{pmatrix} 1\\1\\2\\3\\3\\2\\2\\3\\1\\5\\3\\4\\1 \end{pmatrix} + 0.8^{4}\begin{pmatrix} 1\\2\\3\\2\\1\\3\\3\\2\\2\\2\\4\\1\\2 \end{pmatrix} + 0.8^{5}\begin{pmatrix} 2\\3\\2\\1\\0\\2\\2\\1\\3\\0\\1\\2\\3 \end{pmatrix} + 0.8^{6}\begin{pmatrix} 3\\2\\1\\0\\0\\1\\1\\0\\2\\0\\0\\1\\2 \end{pmatrix} + 0.8^{7}\begin{pmatrix} 2\\1\\0\\0\\0\\0\\0\\0\\1\\0\\0\\0\\0 \end{pmatrix} \\
&\quad + 0.8^{8}\begin{pmatrix} 1\\0\\0\\0\\0\\0\\0\\0\\0\\0\\0\\0\\0 \end{pmatrix} + 0.8^{9}\begin{pmatrix} 0\\0\\0\\0\\0\\0\\0\\0\\0\\0\\0\\0\\0 \end{pmatrix}
\end{aligned}
$$

The resulting neighbourhood centralities are:

$$
\delta(0.8) = \begin{pmatrix} 4.39 & 5.23 & 6.05 & 7.00 & 7.55 & 6.05 & 6.21 & 7.00 & 5.13 & 6.9 & 6.38 & 6.26 & 5.56 \end{pmatrix} \tag{7}
$$

We see from eq. (7) that node 5 has the highest neighbourhood centrality, followed by nodes 4 and 8. The least central node is 1, and the reason, according to Figure A1, is that node 1 has only one first-order neighbour and the contribution from most of the other nodes to the centrality of node is highly discounted since they are multiple steps away.

Figure A1: An example of a directed network with the corresponding adjacency matrix G. Node 5 has the highest neighbourhood centrality score of 7.55, followed by nodes 4 and 8 with a centrality score of 7. The node with the least neighbourhood centrality is node 1, with a centrality score of 4.39.



$$G = \begin{pmatrix} 0&1&0&0&0&0&0&0&0&0&0&0&0 \\ 1&0&1&0&0&0&0&0&0&0&0&0&0 \\ 0&1&0&1&0&0&0&0&0&0&0&0&0 \\ 1&0&1&0&1&0&0&0&0&0&0&0&0 \\ 0&0&0&1&0&1&0&0&0&1&0&0&0 \\ 0&0&0&0&0&0&1&1&0&0&0&0&0 \\ 0&0&0&0&0&1&0&1&1&0&0&0&0 \\ 0&0&0&0&1&1&1&0&0&0&0&0&0 \\ 0&0&0&0&0&0&1&0&0&0&0&0&0 \\ 0&0&0&0&1&0&0&0&0&0&1&0&0 \\ 0&0&0&0&0&0&0&0&0&1&0&1&0 \\ 0&0&0&0&0&0&0&0&0&0&1&0&1 \\ 0&0&0&0&0&0&1&0&0&0&0&0&0 \end{pmatrix}$$