Of stars and galaxies — Co-authorship network and research

Hu, Xiaoli; Li, Oliver Zhen; Pei, Sha

Published in:
China Journal of Accounting Research

Published: 01/03/2020

Document Version:
Final Published version, also known as Publisher’s PDF, Publisher’s Final version or Version of Record

License:
CC BY-NC-ND

Publication record in CityU Scholars:
Go to record

Published version (DOI):
10.1016/j.cjar.2019.09.002

Publication details:
Hu, X., Li, O. Z., & Pei, S. (2020). Of stars and galaxies — Co-authorship network and research. *China Journal of Accounting Research*, *13*(1), 1-30. https://doi.org/10.1016/j.cjar.2019.09.002

China Journal of Accounting Research 13 (2020) 1–30

Contents lists available at ScienceDirect

China Journal of Accounting Research

journal homepage: www.elsevier.com/locate/cjar

Of stars and galaxies – Co-authorship network and research[☆]

Xiaoli Hu[a], Oliver Zhen Li[b,c,*], Sha Pei[b]

[a] City University of Hong Kong, Hong Kong, China
[b] Shanghai Lixin University of Accounting and Finance, China
[c] National University of Singapore, Singapore

ARTICLE INFO

Article history:
Received 17 September 2019
Accepted 26 September 2019
Available online 29 November 2019

Keywords:
Network centrality
Co-authorship
Research output
Accounting research community

ABSTRACT

We examine the association between network centrality and research using the accounting research community setting. We establish co-authorship network using papers published in the five top accounting journals from 1980 to 2016. We find that the co-authorship network in accounting is a ‘‘small world" with some most connected authors playing a key role in connecting others. We use machine learning to label published papers with multiple topics and find patterns in topics over time. More importantly, we find that co-authorship network centrality is positively associated with future research productivity and topic innovation and that the impact of centrality on productivity is higher with more senior authors. Further, centrality of an author's co-authors also has an incrementally positive impact. We conclude that network centrality positively influences research output.

1. Introduction

A growing literature examines the economic consequences of network centrality based on personal connections among firm executives, board members, etc. For example, Faleye et al. (2014), El-Khatib et al. (2015)

[☆] We thank seminar participants at the Chinese University of Hong Kong, Chinese University of Hong Kong (Shenzhen), Donghua University, Nanjing University, Renmin University of China, Shanghai University of Finance and Economics, University of Tennessee, University of Georgia, University of Hawaii, National University of Singapore, Chongqing University, Nanjing Audit University, Shanghai Jiaotong University, and Shanghai Lixin University of Accounting and Finance.

[*] Corresponding author at: School of Accounting, 817 Xulun Building, Shanghai Lixin University of Accounting and Finance, 2800 Wenxiang Road, Songjiang District, Shanghai 201620, China.

E-mail address: bizzhenl@nus.edu.sg (O.Z. Li).

https://doi.org/10.1016/j.cjar.2019.09.002
1755-3091/© 2019 Sun Yat-sen University. Production and hosting by Elsevier B.V.

and Larcker et al. (2013) focus on the impact of network centrality on firm performance. However, they reach different conclusions. We study another important form of network, the co-authorship network among research scholars, and examine the impact of network centrality on individual researcher's output.

Network plays an especially important role in research. Collaboration in research activities is common and increasing among firms, organizations, and individuals (Becker and Dietz, 2004; Cowan et al., 2007; de Faria et al., 2010). The co-authorship network has expanded dramatically overtime. Further, researchers rely much more, compared with other work forces, on interactions with collaborators (Allen, 1971, 1977; Allen et al., 2007). Accordingly, the research community offers a unique and excellent setting for examining the association between network centrality and output.

We study the association between network centrality and output by employing a novel dataset on the co-authorship network constructed based on the publication records of accounting researchers. Specifically, using 5895 papers published in the top five accounting journals (*Journal of Accounting Research – JAR*; *The Accounting Review – TAR*; *Journal of Accounting and Economics – JAE*; *Contemporary Accounting Research – CAR*; and *Review of Accounting Studies – RAS*) from 1980 to 2016, we establish a co-authorship network and examine its property. We apply machine learning to label each paper with multiple topic tags, which allows us to more objectively depict the development of research ideas in the accounting literature in a large-sample setting. We then estimate the impact of individual authors' centrality in the co-authorship network on their future research output.

To establish a co-authorship network, we treat every author as a node in the network and define two authors as linked if they have a co-authorship relationship through published papers. Goyal et al. (2006) define a network as having a "small world" property if it satisfies four features. First, the number of nodes (authors) is very large relative to the number of links (co-author relationships). Second, there exists a giant component that covers a large proportion of the population.[1] Third, the average number of steps needed to connect any two nodes in the network is small. Fourth, within the network, the clustering coefficient which measures the overlap of co-authorship is high.[2] Our results show that the co-authorship network in accounting satisfies all four features. We further check the role played by the most connected authors who have high co-authorship compared with the average co-authorship of the population. These authors generally have more publications and a higher proportion of co-authored papers. More importantly, the clustering coefficient of the most connected authors is low relative to the network average. This suggests that, while these authors collaborate with many co-authors, their co-authors generally do not work with each other. Therefore, the most connected authors occupy an important position in sustaining the network.

Our main purpose is to examine the impact of co-authorship-based network centrality on individuals' research output, as measured by productivity and topic innovation. Centrality reflects the importance of an individual author in the network. We use three measures to capture centrality. The first measure is *Degree*, which is the number of co-authors for an author in the network. The second measure is *Closeness* (Sabidussi, 1966), which captures the inverse of the overall distance of an author to other authors. The third measure is *Betweenness* (Freeman, 1977), which captures the extent to which the shortest path between two authors goes through a given author. The higher are the values of these three measures, the more central an author is in the network. We find that an author's network centrality is positively associated with an author's future productivity and topic innovation.

As the three centrality measures are calculated based on the co-authorship network formed via an author's past publications, this can cause two potential endogeneity issues. First, if both an author's research output and network centrality are affected by common omitted variables, the association between co-authorship network centrality and research output can be biased. Second, it is possible that productive authors can attract more co-authors and hence gain centrality in the network. We use an instrumental variable approach to address endogeneity and the general tone of our findings does not change.

We execute several further analyses on the association between network features and research output. First, author seniority enhances the effect of centrality on research productivity but not on topic innovation. Second,

[1] A component is a sub-group of nodes in the network within which any two nodes can be connected through one or several steps.

[2] The clustering coefficient measures the extent to which an author's co-authors are also co-authors with each other. The higher is this coefficient, the more overlapping is co-authorship in the network.

while isolated authors have lower research productivity, they have a similar level of topic innovation compared with other authors. Third, we show that centrality of an author's co-authors also has an incrementally positive impact on research output. Overall, we provide evidence that network centrality affects research output.

We make several contributions. First, we contribute to the literature on the economic consequences of network centrality. Earlier studies have examined network centrality based on personal connections among board members/executives, and the impact of network centrality on firm performance, innovation and the cost of debt (Larcker et al., 2013; Chullun et al., 2014; Akbas et al., 2016; Faleye et al., 2014; El-Khatib et al., 2015). However, whether the effect of network centrality is positive or negative is far from being clear. We recognize that networking and collaboration are especially important in the academic research community. Using the accounting research community as a setting, we show that co-authorship-based network centrality positively influences research output. Ahuja et al. (2003) examine how individual centrality affects performance in a Virtual R&D group. Our study differs from them in several ways. First, individuals in their study all belong to a formal R&D group and thus have official access to each other, while in our study, most authors are not affiliated with the same organization. Second, they construct networks based on communication (email-messages) while we focus on co-authorship based network. Third, due to the nature of the individual network, Ahuja et al. (2003) focus on how individual centrality mediates the effects of individuals' functional, status and communication role on their performance. We, on the other hand, are interested in how centrality in the co-authorship network directly affects authors' research output.

Second, we contribute to an analysis of the accounting research community. While researchers can establish various social connections such as work affiliations, doctoral programs (Lohmann and Eulerich, 2017) or paper citations (Bonner et al., 2012), co-authorship is a particularly important form of social connection. Co-authorship has gained increased popularity as the communication cost has decreased substantially and research projects have become more challenging and complex. It is a long-lasting relationship that involves intense collaborations and risk-sharing. Researchers have examined the co-authorship network in other disciplines (Goyal et al., 2006; Goldenberg et al., 2010). We describe the co-authorship network in accounting and provide further insights into the evolution and impact of social networks in the research community.

Third, we apply a new research topic classification. Prior studies in this field classify accounting papers into a limited number of subjective topics (Oler et al., 2010). The Latent Dirichlet Allocation (LDA) model we employ generates multiple topic labels on an objective basis. Based on the LDA topic labels, we construct an author-paper level topic innovation measure. The topic innovation measure has two advantages. First, as the LDA topic modelling allows us to divide the accounting literature into a large number of topics, we are able to examine topic innovation for a large sample. Second, the LDA topic modelling is an objective algorithm that does not require discipline-specific information. Our measure can thus be easily applied to other disciplines to evaluate paper-level topic innovation. In this sense, our study not only contributes to an analysis of the accounting literature, but also provides a useful research output measure for analyses of literatures in other disciplines.

Finally, our research has "policy implications". Predicting research output is of great importance to universities in their recruiting and promotion decision-making process. Our findings suggest that co-authorship network centrality helps predict research output in addition to researchers' past number of publications, thus assisting universities' evaluation process.

This paper proceeds as follows. Section 2 reviews literature and formulate our hypothesis. Section 3 discusses sample formation. Section 4 describes the establishment of the co-authorship network and examines its property. Section 5 discusses methodologies and results of the paper topic analysis. Section 6 reports results on the impact of network centrality on research output. Section 7 summarizes and concludes.

2. Literature and hypothesis development

2.1. Literature on network centrality and hypothesis

Social networks provide channels for the flow of influence, support, information and other valuable resources among people and organizations (Larcker et al., 2013). Network centrality describes the position of individuals or organizations in a network. The higher the centrality, the more important a person or an organization is in a network, and the easier it becomes for them to access resources in the network. Accordingly, several studies find that network centrality has positive effects on people and organizations. Faleye, Kovacs, and Venkateswaran (2014) use social centrality to measure CEO social connections. They argue that strong social connections provide CEOs with an information advantage to explore and utilize innovative ideas as well as high job security in the labour market, and thus reduce CEOs' risk aversion. Accordingly, they document that better-connected CEOs invest more in research and development and that their firms obtain more high quality patents. Larcker et al. (2013) argue that high board centrality can facilitate information and resource exchanges among firms, bring social capital to them and foster collaborations. They find that board centrality improves stock returns and return on assets. Chullun et al. (2014) suggest that high board centrality can enhance investor recognition, help firms build closer ties with financial institutions, and reduce information asymmetry by increasing firm visibility and reputation. They demonstrate empirically that high board centrality expands firms' access to external capital and reduces their cost of debt.

Network plays an especially important role in the research community. Increased complexity in knowledge creation and innovation has led to a tremendous growth in collaboration among firms, organizations and individuals in research activities (Becker and Dietz, 2004; de Faria et al., 2010). In the academic world, the proportion of co-authored papers has increased significantly, in almost all disciplines, over the past several decades. As a result, the co-authorship network has expanded dramatically overtime. Further, it has been documented that researchers rely five times more, compared with other work forces, on interactions with collaborators in their work (Allen, 1971, 1977; Allen et al., 2007). Therefore, the research community offers us a unique setting to examine the association between network centrality and output. In fact, as researchers, and especially academic researchers, rely more on collaboration, a positive association between network centrality and output is more likely to be found in the research community than in other business settings. Accordingly, we propose the following hypothesis:

Hypothesis. *There is a positive association between co-authorship-based network centrality and research output.*

Several features of the academic research community also make it an ideal setting for examining the association between network centrality and output. First, we can clearly identify individuals as well as their co-authorship links, which are important for establishing the social network. Second, due to the common practice of publishing university faculty members' CVs online, we can collect detailed personal information of researchers. Third, a publication-based evaluation system used by universities allows for better measurement of research output. While we normally use patents as a measure of firms' innovation output, many firms choose to keep some of their technologies as business secrets. As such, patent-based measures cannot capture the full picture of firms' innovation output and they are also endogenous to firms' operational and business decisions. Academic researchers' published papers less ambiguously reflect their research output. Fourth, data and machine learning technology allow us to construct a novel research output measure, topic innovation, which captures the extent to which academic researchers push their boundary of knowledge and explore new topics. Finally, the academic community is a proper setting to examine our research question as university researchers often have the freedom to choose topics that they are interested in and are less prone to commercial biases. Researchers in commercial institutions often have to conform to their employers' overall business strategies and thus face more restrictions in determining the topics and the interpretation of their results.

Of course, network centrality can also be associated with something negative. For example, El-Khatib et al. (2015) argue that CEOs with high centrality are more powerful and can exert a greater influence on their boards, which can potentially mitigate the effect of internal governance on CEOs and hurt shareholder value. They document a negative impact of CEO network centrality on merger performance. In the case of academic researchers, being central in a network can excessively consume their time, energy, and attention, and thus reduces their research output. This possibility adds tension to our prediction.

2.2. Patterns in research publications in accounting and other disciplines

A strand of literature examines patterns in academic publications. Hasselback et al. (2000) examine both the quantity and quality of publications of accounting scholars graduated from 1971 to 1993. Following this work, Glover et al. (2006) and Glover et al. (2012) examine publication records of accounting scholars in the top 75 schools when they were promoted to associate or full professors from 1995 to 2009. Oler et al. (2010), by classifying citations according to disciplines, show that finance and economics make a growing contribution to the origination of ideas in accounting research. They further divide published articles into six topic categories (financial accounting, managerial accounting, auditing, tax, governance and other topics) and seven methodology categories (archival, experimental, field study, review, survey, theoretical and normative) and show changes in research topics and methodologies over forty eight years.

Some recent studies have started to investigate social networks and their impact on the accounting research community. Bonner et al. (2012) examine social structure through which accounting research ideas are communicated based on citations among authors. Lohmann and Eulerich (2017) identify and describe institutional networks based on work affiliations as well as doctoral programs for papers published in *The Accounting Review*. They find that while the work affiliation network has become more diverse overtime, the network based on Ph.D granting institutions is still concentrated in a relatively small group of universities.

Researchers in other disciplines have examined certain properties and impacts of the co-authorship network in their fields. Goldenberg et al. (2010) establish collaboration networks in marketing over forty years. Goyal et al. (2006) study social distance among economists based on the co-authorship network from 1970 to 2000 and identify "stars" in networks. Studies in the economics literature have also examined how co-authorship network properties affect individual authors' research productivity (Hollis, 2001, Medoff, 2003, Ductor, Fafchamps, Goyal and van der Leij, 2014; Ductor, 2015). We examine the property and development of the co-authorship network in the accounting research community. Further, we develop a research output measure, topic innovation, in addition to the traditional measure of research productivity.

3. Data

We focus on five top accounting journals, *JAR*, *TAR*, *JAE*, *CAR* and *RAS* from1980 to 2016.[3] We collect information on *JAE* papers from ScienceDirect, information on *JAR* and *TAR* papers from Ebscohost, information on *CAR* papers from ProQuest and information on *RAS* papers from Springer. For each paper, we obtain its title, author name(s), publication time and the abstract.

A key step in constructing the co-authorship network is identifying all unique authors and their publications in these five journals. While the data we collect contain author name(s) of each paper, an author's name can be presented differently in different papers and journals. To distinguish authors, we compare their last names and the initials of all their first names. For names that we are suspicious of duplications or errors, we manually check the original papers or authors' resumes. As a result, we identify 3628 unique authors. We also collect Ph.D graduation information for each author from the Brigham Young University (BYU) accounting researcher ranking database. For authors with missing graduation information in the BYU database, we manually collect this information from their resumes.

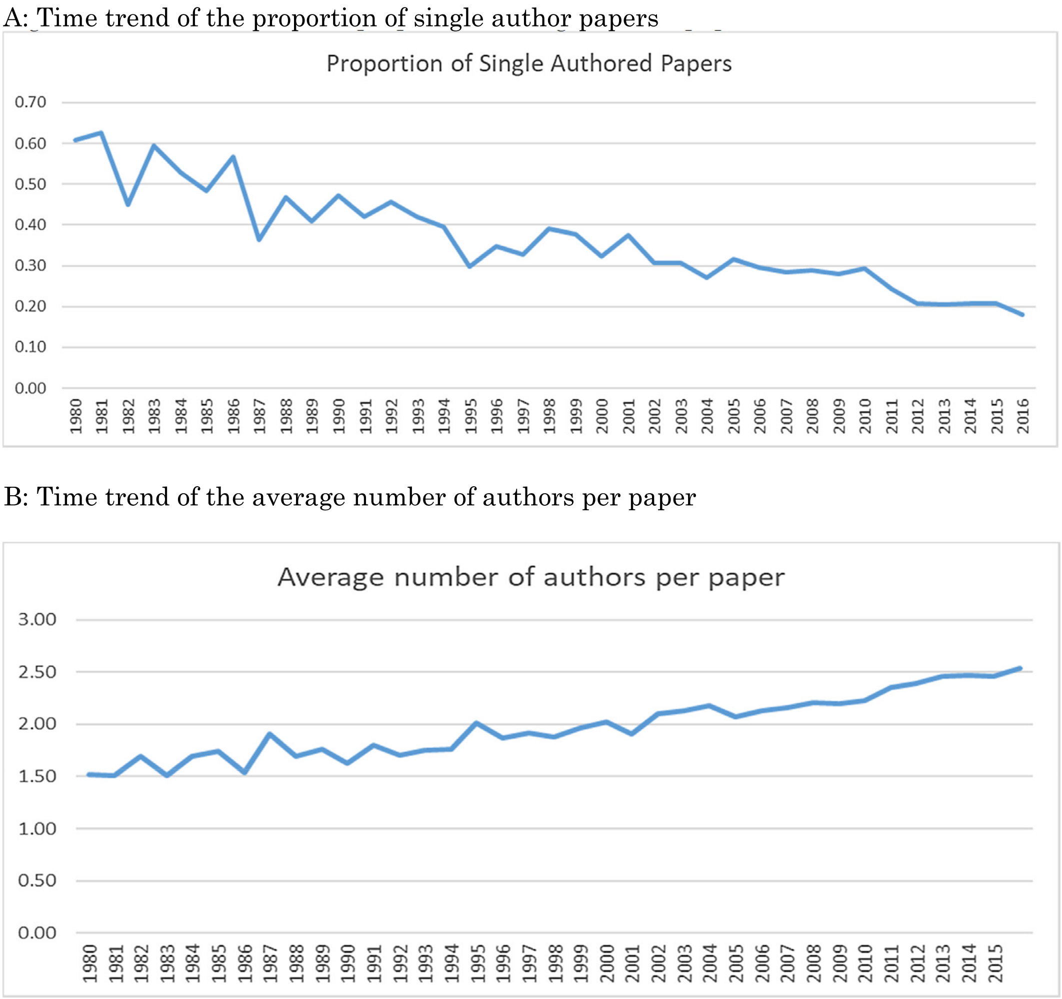
We calculate the percentage of single-author papers and the average number of authors per paper over time and show their time trends in Fig. 1. Fig. 1-A is the time trend of the proportion of single-author papers. It declines from above 0.6 in 1980 to below 0.2 in 2016. The average number of authors per paper increases from around 1.5 in 1980 to more than 2.5 in 2016 in Fig. 1-B. Overall, the practice of co-authorship has become increasingly popular over time.

4. Co-Authorship network

4.1. Establishing the Co-Authorship network

We establish co-authorship network in a similar way as in Goyal et al. (2006) and Ductor (2015). Let $G_{t,s}$ denote a co-authorship network from Years $t-(s-1)$ to t. In the network $G_{t,s}$, $N_{t,s} = \{1, 2, \ldots, N\}$ is the set of authors who have publication(s) during the period $t - (s - 1)$ to t. Two authors are linked through co-

[3] For *CAR* the data started in 1984 and for *RAS* the data started in 1996.



A: Time trend of the proportion of single author papers

B: Time trend of the average number of authors per paper

Fig. 1. Time trends of the proportion of single author papers and the number of authors per paper.

authorship between them. Specifically, $g_{i,j}$ is set to one if Authors i and j have published one or more papers together and zero otherwise.

To describe the network $G_{t,s}$, we first define a set of variables. *Total Authors*$_{t,s}$ (N) refers to the number of authors who publish at least one paper during the period $t-(s-1)$ to t. *Isolated Authors*$_{t,s}$ refers to authors that have zero co-authors over the period $t-(s-1)$ to t. A low percentage of isolated authors suggests that many researchers in the network are connected through co-authorship. *Degree* ($d_{i;\ t,s}$) is the number of co-authors for Author i over the period $t-(s-1)$ to t. A high value for *Degree* suggests that an author is central in a network. For the whole network, the average of *Degree* is given by

$$d(Gt,s)=\frac{\sum_{i\in N_{t,s}}d_{i;t,s}}{N}. \tag{1}$$

Following Watts and Strogatz (1998), we devise a *Clustering Coefficient* ($CL_{i;\ t,s}$) to measure the percentage of Author i's co-authors who are also co-authors with each other. Its formal definition is

$$CL_{i;t,s}=\frac{\sum_{l\in N_{i,t,s}}\sum_{k\in N_{i,t,s}}g_{l,k}}{d_{i;t,s}(d_{i;t,s}-1)} \tag{2}$$

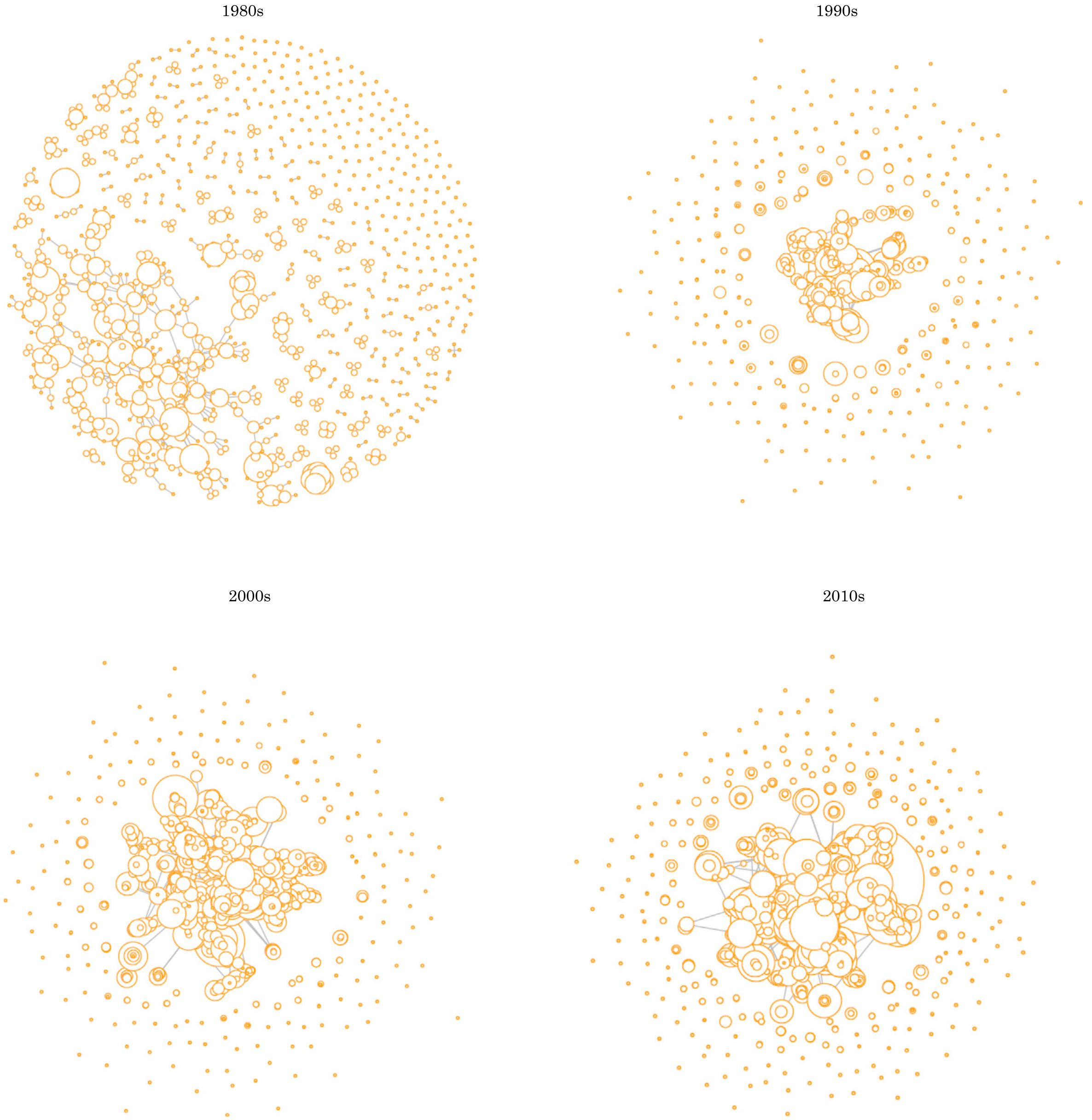


Fig. 2. Networks of co-authorship in top accounting journals from 1980s to 2010s. These figures present networks of co-authors based on publications on the top-five journals from 1980s to 2010s. The size of nodes is proportional to its degree.

where $\mathrm{N}_{\mathrm{i,t,s}}$ is the set of co-authors for Author *i*. $CL_{i;\ t,s}$ is available for authors with $d_{i;t,s} \geq 2$. To measure the overall *Clustering Coefficient* of a whole network, we use the weighted average of $CL_{i;\ t,s}$ that is

$$CL(Gt,s) = \sum_{i \in N'_{t,s}} \frac{d_{i;t,s}(d_{i;t,s}-1)}{\sum_{j \in N'_{t,s}} d_{i;t,s}(d_{i;t,s}-1)} CL_{i;t,s} \tag{3}$$

where $\mathrm{N}'_{\mathrm{t,s}}$ is the set of authors where each author's *Degree* is larger than or equal to two.

If two authors have co-authorship or if there is a set of distinct intermediate co-authors that link them, they are connected by a path. For each path, we count the number of steps it takes to connect two authors. Multiple paths can exist between any two authors. The length of the shortest path is defined as the distance between Authors *i* and *j, dist*(*i,j; $G_{t,s}$*). Thus, the distance captures the most efficient way to connect two authors.

A component in $G_{t,s}$ is a sub-network where each pair of authors can be connected, directly or indirectly. A *Giant Component* (*GC(*$G_{t,s}$)) is the largest component that contains a significant number of authors and it dominates other components. We calculate the average distance of the giant component as below

$$dist(GC(Gt,s)) = \frac{\sum_{\in N''_{t,s}} \sum_{j \in N''_{t,s}} dist(i,j;G_{t,s})}{N''(N''-1)} \tag{4}$$

where $\mathrm{N}''_{t,s}$ is the set of authors that belong to *GC(*$G_{t,s}$) and N'' is the number of authors in *GC(*$G_{t,s}$). *Diameter* of *GC(*$G_{t,s}$) is the longest length of the shortest distance in the *Giant Component*. It captures the largest number of steps to connect two authors in this component. These two measures capture the overall connectedness of the *Giant Component*.

Goyal et al. (2006) define a network as having a “small world” property if it satisfies four features. First, the number of authors is significantly larger than the average *Degree*: $\mathrm{N} \gg \mathrm{d(Gt,s)}$. On average, each author does not have many co-authors compared with the size of the network. Second, the network needs to have a giant component that contains a large number of authors in the network. Although authors do not have many direct connections with each other through co-authorship, many of them can still be linked indirectly. Third, the average distance between authors in the giant component is short, e.g., $\mathrm{dist(Gt,s)}$ is of order ln(*N*), where ln(*N*) represents the average distance of a typical random network. That is, it does not need too many steps to connect any two authors. Fourth, clustering is high, e.g., $CL(Gt,s) \gg d(Gt,s)/N$, where $\mathrm{d(Gt,s)/N}$ is the typical clustering coefficient value if the network is random. That is, an author’s co-authors have a significant chance of becoming co-authors in a network with a “small-world” property compared with those in a random network.

4.2. Co-Authorship network in five top journals

Table 1 presents statistics for the co-authorship network during the period 1980 to 2016. From 1980s to 2000s, we construct networks in 10-year windows. The network in the 2010s only covers papers from 2010 to 2016.

In Table 1, from 1980s to 2010s, the number of authors in the network has increased from 957 to 1870. However, the percentage of isolated authors has decreased significantly from 19.3% to 5.4%. Thus, co-

Table 1
Descriptive statistics for co-authorship networks.

	1980s	1990s	2000s	2010s
Total Authors	957	1036	1406	1870
Isolated Authors:				
Number	185	121	97	101
Percentage	0.193	0.117	0.069	0.054
Degree				
Mean	1.808	2.174	3.026	3.317
Std	1.789	1.959	2.506	2.816
Clustering coefficient	0.377	0.292	0.297	0.300
Giant Component:				
Size	335	492	1030	1329
Percentage	0.350	0.475	0.733	0.711
Average distance	8.219	7.701	7.098	6.687
Clustering coefficient	0.282	0.262	0.283	0.278
Diameter	21	23	16	17
Second largest component	15	14	13	15

This table presents summary statistics of co-authorship networks from 1980s to 2010s, based on articles published in *Journal of Accounting Economics* (*JAE*), *Journal of Accounting Research* (*JAR*), *The Accounting Review* (*TAR*), *Contemporary Accounting Research* (*CAR*), and *Review of Accounting Studies* (*RAS*).

authorship has become a common practice in the accounting academia. The ease of communication due technological advances, increased difficulty in publishing in top journals and the requirements for different skills, can be reasons for increased collaborations among authors.

We check whether our whole co-authorship network satisfies the four features of the small world property. For the first feature, we examine the average degree of the network $\mathrm{d(Gt,s)}$. Although the average degree has increased from 1.808 in the 1980s to 3.317 in the 2010s, it is small relative to the total number of authors in every period. For instance, in 1990, the average degree is 2.174, which is just 0.21% (2.174/1036) of the total number of authors. Therefore, the first feature of the small world property is satisfied.

Next, we examine the second feature. The size of the largest component is 335 in 1980s, 35.01% of the population. The largest component has grown substantially over time. In 2000s, it is 73.26% of the population with 1030 authors. The percentage in 2010s is smaller, due to the 2010s sample covering a shorter period. The size of the second largest component does not change significantly over time and it is small compared with the largest component. For instance, the second largest component in 2000s contains 13 authors and is only 0.92% (13/1406) of the largest component. Therefore, a giant component exists and has expanded significantly over time. As a result, the second feature is also satisfied.

We then move on to the third feature and examine the average distance of the giant component. The average distance of the giant component is 8.219 in 1980s, 7.701 in 1990s, 7.098 in 2000s, and 6.687 in 2010s. The average distance has been small and has decreased by around 13.17% ((6.687–8.219)/8.219) from 1980s to 2010s. On average, it takes no more than eight steps to connect two authors in the giant component. We compare the average distance of the giant component with ln(N), the typical average distance of a random network. ln(N) is 6.864, 6.943, 7.249 and 7.534 from 1980s to 2010s. The two values are comparable in all periods. Therefore, the third feature of the small-world property is also satisfied.

Finally, we examine the fourth feature. As mentioned earlier, the clustering coefficient measures the extent to which an author's co-authors are also co-authors to themselves. The clustering coefficient of the giant component is 0.282 in 1980s, 0.262 in 1990s, 0.283 in 2000s, and 0.278 in 2010s. We compare the clustering coefficient with $d(Gt,s)/N$, which is the probability of co-authorship formation when the link in the network is randomly assigned. In every period, the clustering coefficient is significantly larger than $d(Gt,s)/N$. For instance, $d(Gt,s)/N$ is 0.002 in 1980s, and the actual clustering coefficient is 188.5 times of this value. The fourth small world feature is satisfied.

Fig. 2 intuitively shows the development of the network and the small world property. In the graph, each node represents an author in the network. Two authors have a co-authorship relationship if there is a link between them. The size of the node is proportional to an author's *Degree*, that is, a large node represents an author with many co-authors. The proportion of nodes with no link (single authors) is becoming smaller. Many nodes are small, suggesting that these authors have a small number of co-authors. However, there is a large group of nodes that can be connected directly or indirectly and this group is becoming significantly larger over time. We can observe a clear expansion of the giant component.

In sum, we conclude that the co-authorship network satisfies the four features of the small world property. While each author does not have many co-authors compared with the number of authors in the network, there exists a giant component where a significant proportion of authors can be linked to others either directly or indirectly. The distance between any two authors is relatively small and has decreased slightly.

4.3. Role of the most connected authors

In Table 2, we examine the distribution of co-authorship in the network and its development over time. Overall, the number of co-authors differs significantly across authors and the gap in *Degree* between the most connected authors and the average has enlarged substantially. For example, *Degree* of the most connected author is 10 and it is 5.531 (10/1.808) times of the average *Degree* in 1980s. In 2010s, this ratio has increased to 11.226 (38/3.385). The most connected authors can also connect with more authors in exactly two steps. *Order 2* is the number of authors that can be connected to Author i in exactly two steps (not in one step). For the most connected author, *Order 2* is 34 in 1980s, 36 in 1990s, 54 in 2000s, and 67 in 2010s. The average *Order 2* is 2.813 in 1980s, 4.290 in 1990s, 8.074 in 2000s, and 10.198 in 2010s. This result suggests that the most connected authors also have a larger pool of potential co-authors.

Table 2
Network statistics for the most connected authors.

Panel A: 1980s

Rank	*Papers*	*%Co-authored*	*Degree*	*Order 2*	*Clustering Coefficient*
1	10	90.0	10	34	0.156
2	12	83.3	10	27	0.089
3	11	72.7	10	20	0.178
Average top 50	6.840	85.0	7.360	10.280	0.299
Average all	2.129	71.0	1.808	2.813	0.377

Panel B: 1990s

Rank	*Papers*	*%Co-authored*	*Degree*	*Order 2*	*Clustering Coefficient*
1	13	84.6	14	36	0.187
2	17	88.2	14	24	0.121
3	8	100	12	32	0.121
Average top 50	8.520	87.0	8.300	19.220	0.157
Average all	2.322	79.2	2.174	4.290	0.292

Panel C: 2000s

Rank	*Papers*	*%Co-authored*	*Degree*	*Order 2*	*Clustering Coefficient*
1	18	100	20	54	0.100
2	13	84.6	18	36	0.072
3	10	100	16	5	0.083
Average top 50	9.440	91.1	11.320	33.040	0.156
Average all	2.544	86.7	3.026	8.074	0.297

Panel D: 2010s

Rank	*Papers*	*%Co-authored*	*Degree*	*Order 2*	*Clustering Coefficient*
1	17	100	38	67	0.058
2	14	92.9	24	94	0.080
3	11	100	18	49	0.144
Average top 50	8.5	96.8	14.380	43.880	0.150
Average all	2.136	90.7	3.385	10.198	0.300

This table presents summary statistics for the most connected accounting scholars in each decade. *Papers* is the number of papers each author has published in that decade. *%Co-authored* is the percentage of papers that are co-authored. *Degree* is the number of co-authors of an author. *Order 2* is the number of authors that can be connected to the author in two steps. *Clustering Coefficient* measures the extent to which an author's co-authors are also co-authors with each other.

Another interesting and important feature of the most connected authors is that they have a low *Clustering Coefficient* than the sample average. For instance, *Clustering Coefficient* for the most connected author is 0.156 in 1980s, 0.187 in 1990s, 0.100 in 2000s, and 0.058 in 2010s. This number is 41.38% (0.156 / 0.377) of the average in 1980s and has declined to 19.33% (0.058 / 0.300) in 2010s. The average *Clustering Coefficient* of the top-50 most connected authors also decreases from 0.299 in 1980s to 0.150 in 2010s. This suggests that while authors with a very high *Degree* have high co-authorship, only a small portion of their co-authors work with each other. Thus, these high *Degree* authors play an important role in sustaining the network. In fact, *Clustering Coefficient* is opposite to our centrality measures introduced below.

5. Measuring publication content

In addition to identifying the co-authorship network, we also examine the contents of papers published in the five accounting journals since 1980. Such information will further our understanding of the development of the accounting literature. We also construct a research topic innovation measure using paper content information for further analysis. We first discuss the method we use to measure paper content.

5.1. Paper topic classification

While one can easily name some research topics in the accounting literature and give examples of papers related to these topics, to objectively label every paper with one or several topics is challenging. Some papers have author specified information related to their contents. One such information is the *JEL* codes. The *JEL* code is a standardized system to classify the economics literature. Usually around three to five *JEL* codes are assigned to a paper. Ductor (2015) use two digit *JEL* codes as topic classifications for papers published in economics journals. There are two problems with *JEL* codes in our sample. First, not all papers have an author specified *JEL* code. Second, *JEL* classification is for the entire economics literature, which is too general and narrow for accounting research. Another author specified information is keywords. However, keywords do not appear to be good topic labels as they involves different information such as measure names, methodologies, research settings, etc. In addition, keywords are not available for many papers in our sample. One available paper topic information system is the paper topic classification from the accounting scholar ranking database maintained by Brigham Young University (BYU). BYU provides a comprehensive list of publications of accounting researchers and labels each paper with several topic tags. However, this classification is also too general. They label a published paper according to research topics and methodologies. Their classification only includes six topics (accounting information system, audit, financial, managerial, tax and other topics) and four research methodologies (analytical, archival, experimental and other methods).

We construct our own classifications. One potential way is to read all the papers and assign topics to these papers. However, it will be subjective when there are no classification standards. To address this problem, we resort to machine learning technology which provides text mining algorithms in topic modelling. Topic modelling is *"a probabilistic framework for the term frequency occurrences in documents in a given corpus*" in machine learning (Grun and Hornik, 2011).[4] In topic models, a corpus is treated as bag of terms.[5] Each corpus can be represented by a vector of term frequencies. This transformation assumes that term orders are negligible, which is referred to the "exchangeability" of terms in the computing science language (Blei et al., 2003). Topic models use the term frequencies of each corpus to generate topic probabilities for each corpus. Different topic models have different assumptions about term distribution and fundamental probabilistic. See Blei and Lafferty (2009) for a review of different topic models and model assumptions.

We use the classic Latent Dirichlet Allocation (LDA) model developed by Blei et al. (2003) to classify sample papers. The LDA algorithm is a Bayesian mixture model which assumes that different topics identified in the model are uncorrelated with each other (Grun and Hornik, 2011). The model assumes that, the collection of term frequencies in each corpus (document) are random variables and represents an infinite mixture distribution. This model considers the exchangeability of both words and documents among topics (Blei et al., 2003). Given a target number of topics N (which is specified by the users) and M documents, the LDA model will generate an M by N probability matrix. The matrix reports the probability of every document to be related to a topic.

Table 3 represents an example of a LDA probability matrix with M documents and N topics. Prob(m, n) refers to the probability of document m to be related to topic n. Note that, for each document m, $\sum_{n=1}^{N}\mathrm{Prob}(m,\ n) = 1$. With a certain cut-off of probability, we can then decide that a paper belongs to several topics with reasonable probabilities. Note that LDA model is an unsupervised topic model. In such a model, there is no training dataset with already known topics. As such, while we can identify which papers are likely to be related to a certain topic, we do not know exactly what that topic is. There are also supervised topic models. In supervised models, one first starts with a training dataset, in which a set of documents is labelled with some topic titles. The supervised model will learn features of topic titles and apply the learned pattern to a new dataset to assign the already specified topic titles to documents in the new dataset. We do not use supervised topic models as the training model will also involve subjective evaluations. First, the pre-determination of topic titles for the training dataset is subjective. Second, even if we can get well specified topic titles for a small training sample, if the training dataset is not representative enough for all potential topics of the whole

[4] In natural language processing, a corpus refers to the text contents of an object (usually a document).

[5] A term refers to a unique word or word root.

Table 3
An example of an LDA topic distribution matrix.

	Topic 1	Topic 2	Topic 3	...	Topic N-1	Topic N
Document 1	Prob(1, 1)	Prob(1, 2)	Prob(1, 3)	...	Prob(1, N-1)	Prob(1, N)
Document 2	Prob(2, 1)	...	...	...	...	...
Document 3	Prob(3, 1)	...	...	...	...	...
Document 4	Prob(4, 1)	...	...	...	...	...
...	...	...	...	...	...	...
Document M−1	Prob(M−1, 1)	...	...	...	...	...
Document M	Prob(M, 1)	...	...	...	...	...

This table represents an example of an LDA probability matrix with *M* documents and *N* topics. Prob(*m, n*) refers to the probability of Document *m* being related to Topic *n*. The sum of the probabilities of all of the topics for a specific document equals one.

literature, the predicted topics will be biased. Since we do not have a systematic and generally accepted topic lists (for example, a list similar to *JEL* code), we cannot ensure that the training dataset is representative. After balancing advantages and disadvantages, we decide that an unsupervised LDA model is a better choice in our setting as it can classify papers according to contents on a relatively objective basis.

5.2. LDA model classification

To apply topic modelling, we need a paper content database to start with. Hall et al. (2008) examine topics in Computational Liguistics from 1978 to 2006 based on whole contents of papers. However, the length of a typical accounting research paper is much longer than a normal article or a typical computing science paper, which will inflate the dimensions of the term frequency vector as well as calculation complexity. Paper contents are affected by both research topics and authors' writing habits. The writing habits can affect the term frequency distribution of each paper, especially when authors assign different proportion of total length of a paper to different parts (such as literature review, empirical discussion, robustness checks, etc). To mitigate these problems, we choose to focus on paper abstracts. A good thing about an abstract is that, it covers the main ideas and results of a paper in limited words and is less affected by writing habits. Not all papers in our sample have an author provided abstract, especially for some early papers. However, the two databases where we collect the paper information from provide their own abstracts for most papers without author specified abstracts. Altogether, we have 5845 papers with available abstracts.

To apply the LDA mode, the first step is to clean the abstract text.[6] Though an abstract is clean relative to the whole body of a paper, some routine transformation is applied to the text. We removed all punctuations and numbers from the text and lowercased all the text. We then remove the stopwords based on the R stopwords vocabulary.[7] We then stem every word in the text so that different formats of a word become the same word root. Then for every term we count how many abstracts contain that term. For terms contained in more than 500 abstracts, we manually check the term list and pick up terms that are common in abstracts but are unlikely to be related to the research topics. We define such terms as customized stopwords and delete them. Finally, we remove terms that appear in less than 3% and that appear in more than 95% of abstracts. The rationale is that, if a term is contained in very few abstracts, it does not capture any common contents and if a term is contained in too many abstracts, it does not contribute to the uniqueness of a document. As a result, these terms are not useful for topic modelling.

We need to specify the number of total topics as an input to the LDA model. We refer to prior literature for benchmarks for the number of topics to classify. Studies on economic journals usually use two digit *JEL* code as topic classification (Ductor, 2015, Fafchamps et al., 2010). There are all together 135 *JEL* two digit codes in the *JEL* classification. Ductor (2015) reported that there are 121 *JEL* two-digit codes from economic journals covered in the EconLit database from 1970 to 2011. Hall et al. (2008) set the topic number parameter to 100 when using the LDA model to examine research topics in the Computational Linguistics literature from 1978

[6] We use the tm package in R to implement the LDA model. See Grun and Hornik (2011) for the description of the package.
[7] Stopwords refers to some commonly used words in a language with no special meanings.

to 2006. Based on these prior studies, we decide that 100 would be a reasonable number of topics. Following Hall et al. (2008), we use Gibbs sampling to estimate the topic probability matrix. This is a commonly implemented LDA model estimation (Griffiths and Steyvers, 2004, Grun and Hornik, 2011).

Panel A, Table 4 shows the distribution of topic probabilities. The 95% percentile of the probability is 3.24%. In other words, if we set the cut-off to be 3.24%, each paper will be attached to 5 topics on average. We finally set the cut-off to be 4%, which results in 20,970 paper-topic pairs, with a paper attached to 3.59 topics on average. This number is comparable to author specified *JEL* codes for *JEL* available papers. Among the 5845 papers, 5796 are labelled with at least one topic with a probability larger than 4%.

Panel B, Table 4 shows the distribution of the number of topics assigned to papers. A majority of the papers are labelled with 2 to 5 topics. Among the 100 topics, the number of related papers also varies. The topic with the fewest papers contains 104 papers and the topic with most papers contains 542 papers.

We calculate some features of topics across years and draw time patterns of these features. Fig. 3-A represents the number of different topics covered each year. We observe that, before 2002, the number of topics is volatile with some years only having as few as around 85 topics. After 2002, the number of topics covered each year becomes stable, ranging from 97 to 100. Fig. 3-B shows the pattern of the number of different topics divided by number of published papers in each year. This figure exhibits a significant drop in the scaled number of topics after 2000. Fig. 3-C shows the Herfindahl Index of topics in each year. The pattern of the Herfindahl Index suggests that research topics in the five journals have become less concentrated in recent years.

Each year, we calculate the number of papers for each topic and identify the top-10 topics accordingly. We then estimate a Probit model which regresses the probability of a topic to be a top-10 topic in Year t on its probabilities in Years $t-5$ to $t-1$. Table 5 reports the auto-regression results. *Top_topic* is an indicator variable which equals 1 if a topic is a top-10 topic in Year t. The sample period is from 1985 to 2015.[8] In Column (1), we estimate the results using papers from all the five journals. We then estimate the model for the five journals separately and report the results in Columns (2) to (6). The coefficients on the lagged *Top_topic* are all significantly positive in the six columns. Fig. 4 presents a visualized form of the magnitudes of these coefficients.

We also examine topic overlap among the five journals. We define $Topic_{i,j,t}$ as an indicator which equals 1 if Topic i is covered in Journal j in Year t and 0 otherwise. We also define $Top\_topic_{i,j,t}$ as an indicator which equals 1 if Topic i is the top-10 topic in Journal j in Year t. For each Topic i-Year t combination, we compute $Topic_{i,JAR,t}$, $Topic_{i,JAE,t}$, $Topic_{i,TAR,t}$, $Topic_{i,CAR,t}$, $Topic_{i,RAS,t}$, $Top\_topic_{i,JAR,t}$, $Top\_topic_{i,JAE,\ t}$, $Top\_topic_{i,TAR,t,}$ $Top\_topic_{i,CAR,t}$ and $Top\_topic_{i,RAS,t}$ respectively. We report correlation coefficients of the ten variables in Table 6. Panel A reports the correlations among $Topic_{i,JAR,t}$, $Topic_{i,JAE,t,}$ $Topic_{i,TAR,t}$, $Topic_{i,CAR,t}$ and $Topic_{i,RAS,t}$. Panel B reports the correlations among $Top\_topic_{i,JAR,t}$, $Top\_topic_{i,JAE,t}$, $Top\_topic_{i,TAR,t}$, $Top\_topic_{i,CAR,t}$ and $Top\_topic_{i,RAS,t}$. We observe that while most of the correlation coefficients are positive and significant, they are lower than 0.2. Results in Table 6 appear to suggest that the five top accounting journals have different topic specializations and/or tastes.

6. Impact of network centrality on research output

6.1. Capturing network centrality

Here, we examine the impact of the co-authorship network on scholars' research output. Specifically, we test how an author's centrality in the network is associated with his research output. Centrality measures the importance of a node (author) in a network. A node of high centrality suggests that it is central or important in the network. Different measures are developed to capture different aspects of centrality. We apply three commonly used centrality measures, *Degree*, *Closeness* and *Betweenness*. As defined earlier, *Degree* is the number of co-authors for Author i in the co-authorship network.

[8] For CAR, the sample starts from 1989 and for RAS the sample starts from 2001. We exclude 2016 data as our database does not cover all papers in 2016.

Table 4
Descriptive statistics of the LDA estimation results.

Panel A: Distribution of topic probabilities

Min	P5	P10	P25	P50	P75	P90	P95	Max
0.10%	0.34%	0.37%	0.43%	0.53%	0.72%	1.94%	3.24%	57.28%

Panel B: Number of topics assigned to a paper

Number of Topics Assigned	Frequency	Percentage (%)
1	302	5.21
2	955	16.48
3	1601	27.62
4	1451	25.03
5	943	16.27
6	401	6.92
7	116	2.00
8	25	0.43
9	2	0.03

This table presents summary statistics of the LDA results. In Panel A, we report the distribution of the topic probability generated by the LDA model. In Panel B, we show the distribution of the numbers of topics with greater than 4% probability assigned to each paper.

Closeness measures the overall closeness of an author to other authors in a network. Its formal definition, based on Sabidussi (1966), is

$$c_{i;t,s} = \frac{N-1}{\sum_{j \in N_{t,s}} d(i,j;G_{t,s})} \tag{5}$$

where $d(i,j;G_{t,s})$ is the distance between Authors i and j. *Closeness* is the inverse of the total distance of Author i from other authors multiplied by $N-1$. When there is no path between Authors i and j, the total number of authors (N) is used instead of $d(i,j;G_{t,s})$. An author with high *Closeness* can easily or quickly reach other authors and hence is more central in the network.

Betweenness measures centrality in terms of an author's role in connecting other authors. We follow Freeman (1977). Specifically, if an author lies on many paths that connect other authors, he is central in the network. The mathematical definition is

$$b_{i;\ t,s} = \sum j \neq k \neq i \in N_{t,s} \frac{\sigma(j,k|i)\sigma(j,k)}{(N-1)(N-2)/2}, \tag{6}$$

where $\sigma(j,k|i)$ is the total number of the shortest paths between Authors j and k that Author i lies on and $\sigma(j,k)$ is the total number of the shortest paths between Authors j and k. Author i with high *Betweenness* is located in a critical position of connecting or communicating with other authors. Therefore, high *Betweenness* means high centrality.

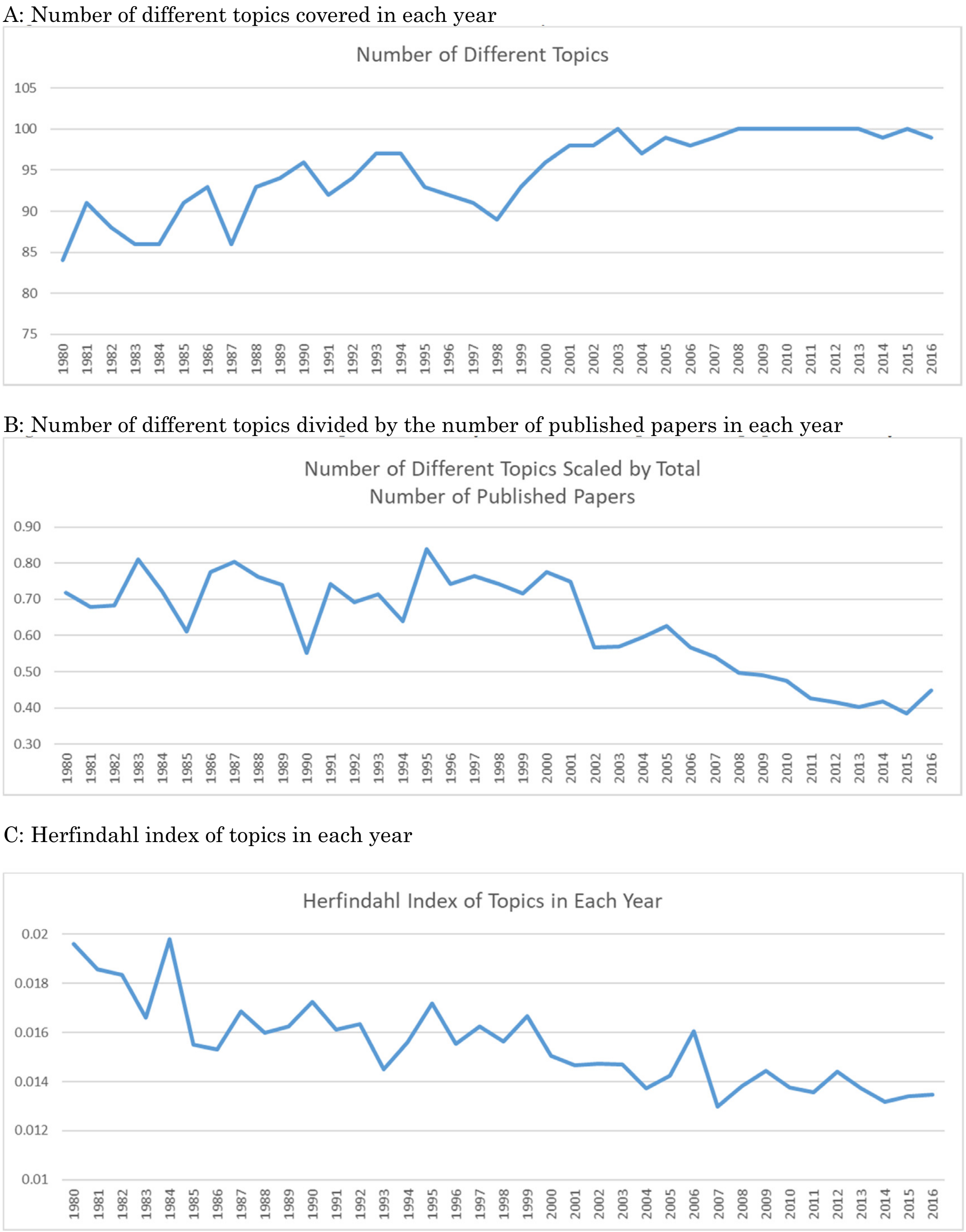
6.2. Author network centrality and research productivity

We apply the following Tobit regression model to examine whether *Centrality* can predict an author's future research productivity:

$$ln(1+q^f_{i;t}) = \beta_0 + \beta_1 Centrality_{i;t,s} + \beta_2 q^p_{i;t,s} + \beta_3 \bar{q}_{1i;t,s} + \beta_4 H_{i;t,s} + D'_{i,t}\omega + University_i + \mu_t + \epsilon_{i;t,s} \tag{7}$$

Our productivity measure $q^f_{i;t}$, is the sum of future three years' productivity, $q^f_{i;t} = q_{i;t+1} + q_{i;t+2} + q_{i;t+3}$, where $q_{i;t}$ is the relative length of papers of Author i in Year t (Ductor, 2015). Specifically,

$$q_{i;t} = \sum_{j=1}^{S} \frac{Pages_j}{Number\ \ of\ \ authors_j} \tag{8}$$



A: Number of different topics covered in each year

B: Number of different topics divided by the number of published papers in each year

C: Herfindahl index of topics in each year

Fig. 3. Topic distribution.

Table 5
Auto-regression of the probability of a topic being a top-10 topic.

	All Five Journals	TAR	JAR	JAE	CAR	RAS
	(1)	(2)	(3)	(4)	(5)	(6)
Variables	*Top_topic*	*Top_topic*	*Top_topic*	*Top_topic*	*Top_topic*	*Top_topic*
Lag_top_topic	0.644	0.670	0.311	0.692	0.553	0.575
	(5.89)***	(6.05)***	(2.72)***	(5.42)***	(4.87)***	(3.49)***
Lag2_top_topic	0.588	0.707	0.786	0.564	0.599	0.774
	(5.26)***	(6.30)***	(7.42)***	(4.48)***	(5.28)***	(5.02)***
Lag3_top_topic	0.880	0.215	0.490	0.836	0.704	0.359
	(8.39)***	(1.74)*	(4.33)***	(6.75)***	(6.46)***	(2.01)**
Lag4_top_topic	0.272	0.610	0.520	0.545	0.437	0.832
	(2.27)**	(5.30)***	(4.91)***	(4.14)***	(3.69)***	(5.31)***
Lag5_top_topic	0.644	0.424	0.555	0.404	0.265	0.533
	(5.88)***	(3.47)***	(4.94)***	(2.97)***	(2.21)**	(3.12)***
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Constant	−1.873	−1.773	−1.633	−1.509	−2.147	−1.495
	(−8.43)***	(−8.97)***	(−8.72)***	(−8.68)***	(−9.53)***	(−8.52)***
N	3100	3100	3100	3100	2700	1500
pseudo R-sq	0.292	0.172	0.141	0.164	0.159	0.182

This table presents the regression results of the following Probit model:

$$Top\_topic = \beta_0 + \beta_1 * Lag\_top\_topic + \beta_2 * Lag2\_top\_topic + \beta_3 * Lag3\_top\_topic + \beta_4 * Lag4\_top\_topic + \beta_5 * Lag5\_top\_topic + Year\ Fixed\ Effects$$

where *Top_topic* is an indicator of whether a topic is among the top-10 topics in year *t*. We regress *Top_topic* on whether the topic is one of the top-10 topics in Years *t* – 5 to *t* – 1.

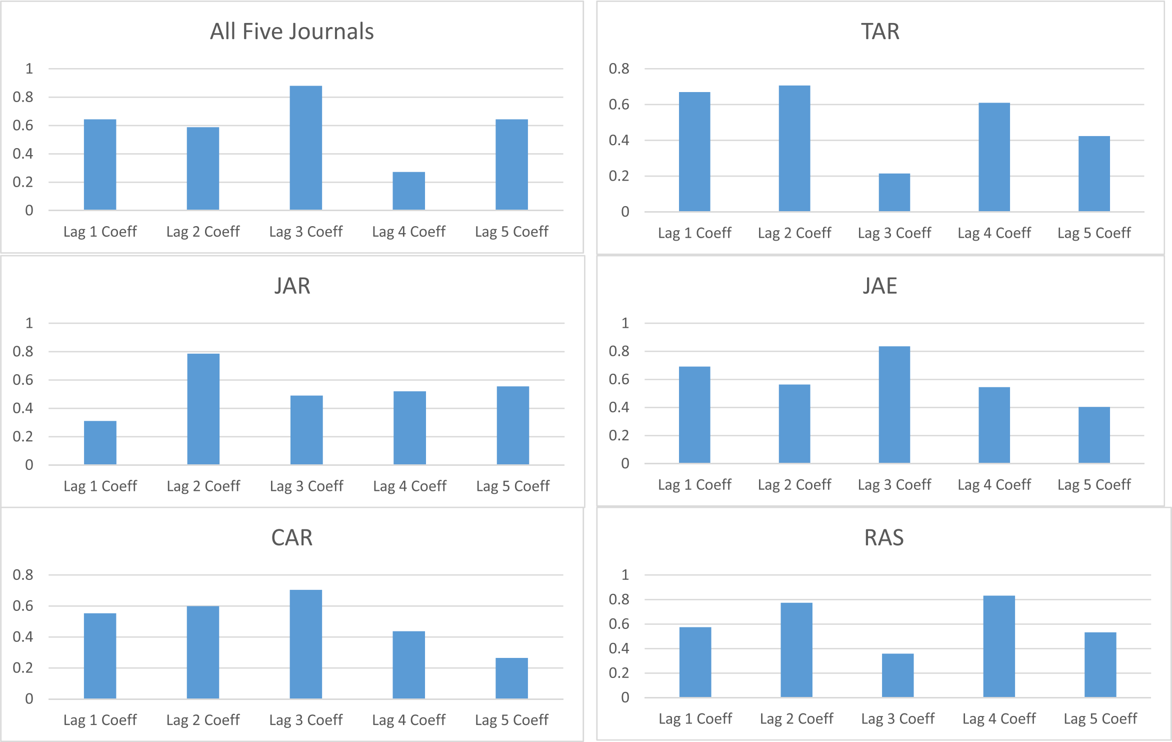


Fig. 4. Time trend of coefficients on lag Top_topics.

Table 6
Correlations for topic coverage among the five journals.

Panel A: Correlations between $Topic_{i,JAR,t}$, $Topic_{i,JAE,t}$, $Topic_{i,TAR,t,}$ $Topic_{i,CAR,t}$, and $Topic_{i,RAS,t}$

	$Topic_{i,TAR,t}$,	$Topic_{i,JAR,t}$	$Topic_{i,JAE,t}$	$Topic_{i,CAR,t}$	$Topic_{i,RAS,t}$
$Topic_{i,TAR,t}$,	1				
$Topic_{i,JAR,t}$	0.1226***	1			
$Topic_{i,JAE,t}$	0.0936***	0.1104***	1		
$Topic_{i,CAR,t}$	0.1617***	0.0771***	0.1076***	1	
$Topic_{i,RAS,t}$	0.1676***	0.1281***	0.1765***	0.1419***	1

Panel B: Correlations between $Top\_topic_{i,TAR,t}$, $Top\_topic_{i,JAR,t}$, $Top\_topic_{i,JAE,t}$, $Top\_topic_{i,CAR,t}$, and $Top\_topic_{i,RAS,t}$

	$Top\_topic_{i,TAR,t}$,	$Top\_topic_{i,JAR,t}$	$Top\_topic_{i,JAE,t}$	$Top\_topic_{i,CAR,t}$	$Top\_topic_{i,RAS,t}$
$Top\_topic_{i,TAR,t}$,	1				
$Top\_topic_{i,JAR,t}$	0.1436***	1			
$Top\_topic_{i,JAE,t}$	0.1034***	0.1330***	1		
$Top\_topic_{i,CAR,t}$	0.1780***	0.1224***	0.0441	1	
$Top\_topic_{i,RAS,t}$	0.0963***	0.1019***	0.0748***	0.0857***	1

This table examines the overlap of topics among the five journals. Panel A presents the correlation coefficients between $Topic_{i,JAR,t}$, $Topic_{i,JAE,t}$, $Topic_{i,TAR,t}$, $Topic_{i,CAR,t}$, and $Topic_{i,RAS,t}$. $Topic_{i,j,t}$ is an indicator which equals 1 if Topic *i* is covered in Journal *j* in Year *t* and 0 otherwise. Panel B represents the correlation between $Top\_topic_{i,TAR,t}$, $Top\_topic_{i,JAR,t}$, $Top\_topic_{i,JAE,t}$. $Top\_topic_{i,CAR,t}$ and $Top\_topic_{i,RAS,t}$. $Top\_topic_{i,j,t}$ is an indicator which equals 1 if Topic *i* is a top-10 topic in Journal *j* in Year *t* and 0 otherwise. *** indicates significance at the 0.01 level.

where *S* is the total number of papers in Year *t* of Author *i*, $Pages_j$ is the number of pages of Paper *j* divided by the average number of pages of papers published during the same year, and $Number\ \ of\ \ authors_j$ is the number of authors of Paper *j*.[9] If there is no paper published in Year *t*, $q_{i;t}$ is set to zero. We use $ln(1+q^f_{i;t})$ to reduce the impact of extreme values. Due to infrequent publications in top journals, $q^f_{i;t}$ has many zero values. We use a Tobit regression model to address the truncation problem.

The main variable of interest is $Centrality_{i;t,s}$, and it includes *Degree, Closeness,* and *Betweenness,* respectively. We calculate centrality measures using 5-year co-authorship network. Therefore, *s* equals 5. We predict that central authors are more productive in the future as they have access to more ideas and have more opportunities for collaboration.

We also control for a set of variables that can affect research productivity. All time dependent control variables are estimated for the same period as the co-authorship network. Past productivity of Author *i* is $q^p_{i;t,s} = q_{i;t} + q_{i;t-1} + \cdots + q_{i;t-(s-1)}$. Authors who are productive in the past are likely to be productive in the future as past productivity can be an indicator of an author's skill and his number of on-going projects.

The average co-author's productivity, $\bar{q}_{1i;t,s}$, over the same period is

$$\bar{q}_{1i;t,s} = \frac{\sum_{j\in N_{i,t,s}} q_{jt}^{-i}}{N_i} \tag{9}$$

where $\sum_{j\in N_i(G_{t,s})} q_{jt}^{-i}$ is the sum of productivity of all co-authors of Author *i*, excluding papers co-authored with Author *i*. N_i is the number of co-authors of Author *i* between $t-(s-1)$ and *t*. $\bar{q}_{1i;t,s}$ is set to zero when Author *i* has no co-author during this period. The effect of co-authors' past productivity is *ex ante* ambiguous. Working with productive co-authors can generate a positive impact as they likely have high quality ideas and are experienced. However, productive co-authors can also have a negative effect as they are busy and thus have limited time for each project.

The degree of specialization, $H_{i;t,s}$, is captured by the Herfindahl index

[9] We also use an alternative productivity measure which replaces $Page_j$ with $Paper_j$, where $Paper_j$ equals one divided by the total number of papers published in the same journal during the same year. Our results are robust to this alternative measure of research productivity.

Table 7
Summary statistics of key regression variables.

Variables	*N*	Mean	Std	25%	Median	75%
Productivity	23,216	0.249	0.340	0.000	0.000	0.443
Past Productivity	23,216	0.606	0.351	0.328	0.501	0.812
Avg. co-authors' past productivity	23,216	0.642	0.760	0.000	0.400	1.023
Degree of specialization (HHI)	23,216	0.250	0.174	0.140	0.200	0.333
Degree	23,216	2.270	1.865	1.000	2.000	3.000
Closeness	23,216	0.002	0.001	0.001	0.002	0.002
Betweenness	23,216	0.001	0.004	0.000	0.000	0.000
Senior	23,216	0.382	0.486	0.000	0.000	1.000
Senior_degree	23,216	0.974	1.745	0.000	0.000	2.000
Senior_closeness	23,216	0.001	0.001	0.000	0.000	0.001
Senior_betweenness	23,216	0.001	0.003	0.000	0.000	0.000
Dif_degree	23,216	0.718	2.079	0.000	0.000	1.500
Dif_closeness	23,216	0.000	0.000	0.000	0.000	0.000
Dif_Betweenness	23,216	0.001	0.005	0.000	0.000	0.001
Percentage of New Topics	8895	0.861	0.209	0.750	1.000	1.000
Percentage of New Co-authors	8895	0.437	0.265	0.333	0.500	0.667
New Co-author's New Topics	8895	9.391	12.238	0.000	4.000	15.000
Gap	8895	1.989	1.172	1.000	2.000	3.000

This table reports summary statistics of variables used in the regression analyses. *Productivity* is the 3-year future productivity of individual authors. *Past Productivity* is the productivity of individual authors during the period when centrality measures are calculated. *Avg. co-authors' past productivity* is the average of past productivity of all the co-authors of an individual author during the period when centrality measures are calculated. *Degree of specialization* is the Herfindahl Index of paper topics for papers published during the period when centrality measures are calculated. *Degree*, *Closeness* and *Betweenness* are the three measures of centrality based on the 5-year network. *Percentage of New Topics* measures the topic innovation in a paper. *Percentage of New Co-authors* measures the percentage of new authors in a paper relative to the most recent paper. *New Co-author's New Topics* is the number of new topics the new co-authors in a paper has done before. *Senior* is an indicator that is equal to one if an author's career seniority is higher than 17 years, and zero otherwise. *Senior_degree, Senior_closeness* and *Senior_betweenness* are interaction terms of *Senior* and the three centrality measures. *Dif_degree* is an author's co-authors' average *Degree* relative to the author's *Degree. Dif_closeness* is an author's co-authors' average *Closeness* relative to the author's *Closeness. Dif_Betweenness* is an author's co-authors' average *Betweenness* relative to the author's *Betweenness. Gap* measures the time gap between the current year paper and the most recent paper when *Percentage of New Topics* is calculated.

$$H_{i;t,s} = \sum_{f=1}^{F} \left(\frac{n_{i;t,s}^{f}}{n_{i;t,s}} \right)^2 \tag{10}$$

where $n_{i;t,s}^{f}$ is the number of articles published between $t-(s-1)$ to t on Topic f, and $n_{i;t,s} = \sum_{f=1}^{F} n_{i;t,s}^{f}$ is the total number of papers of Author i between $t-(s-1)$ and t. The effect of $H_{i;t,s}$ is also *ex ante* unclear. Finally, we control for career seniority ($D_{i,t}$), year (μ_t), and Ph.D granting university ($University_i$) fixed effects. $D_{i,t}$ captures the impact of experience on research output. To construct $D_{i,t}$, we first define $t_{i,0}$ as the PhD graduation year minus five and career seniority as $c_{i,t} = t_i - t_{i,0}$.[10] $D_{i,t}$ are indicator variables for $c_{i,t}$. μ_t are year indicator variables and help control for a potential time trend in author research productivity. University$_i$, controls for the Ph.D granting university effect.

Summary statistics for dependent and independent variables are reported in Table 7. Regression results of Model (7) are presented in Table 8. Continuous variables in this model and subsequent models are winsorized at the 1st and 99th percentiles. We calculate the centrality measures using the 5-year network from Years $t-4$ to t. The coefficients on the three centrality measures are positive and significant. Central authors in the network have high future productivity. The coefficient on *Degree* is 0.026 ($t = 22.73$). Authors with many co-authors generate high productivity in the future. These authors have better access to ideas and can engage in more research projects. The coefficient on *Closeness* is 73.449 ($t = 24.55$). Authors that are close to other authors can easily connect with them and hence are more productive in the future. The coefficient on *Between-*

[10] We subtract five from graduation year as some authors have published papers before Ph.D graduation. Still, several authors have publications before the start of PhD, and we use the first year of publication as $t_{i,0}$.

Table 8
Network centrality and research productivity.

Variables	(1)	(2)	(3)
	Productivity	*Productivity*	*Productivity*
Degree	0.026 (22.73)***		
Closeness		73.449 (24.55)***	
Betweenness			4.186 (13.65)***
Past productivity ($q^{p}_{i;t,s}$)	0.569 (100.14)***	0.606 (105.41)***	0.595 (106.27)***
Avg. co-authors' past productivity ($\bar{q}_{1i;t,s}$)	0.049 (18.15)***	0.038 (13.37)***	0.053 (19.71)***
Degree of specialization ($H_{i;t,s}$)	−0.113 (−9.55)***	−0.156 (−12.52)***	−0.166 (−13.65)***
Year fixed effects	Yes	Yes	Yes
Career seniority fixed effects	Yes	Yes	Yes
Ph.D Granting university fixed effects	Yes	Yes	Yes
N	23,216	23,216	23,216
Pseudo R-sq	0.180	0.179	0.178

This table presents results on the effect of author network centrality on research productivity. We estimate the following model:

$$Productivity_{i;t} = \beta_0 + \beta_1 Centrality_{i;t,s} + \beta_2 Past\ Productivity_{i;t,s} + \beta_3 Avg.coauthors'\ past\ productivity_{i;t,s}$$
$$+\beta_4 Degree\ of\ specialization_{i;t,s} + Year\ Fixed\ Effects + Career\ seniority\ fixed\ effects$$
$$+Ph.D\ Granting\ university\ fixed\ effects,$$

where *Productivity*$_{i;t}$ is the 3-year future productivity of Author *i* in Year *t*. *Centrality*$_{i;t,s}$ represents the centrality measures for Author *i* during the period of Years $t-4$ to *t*. We use *Degree, Closeness* and *Betweenness* to measure *Centrality* respectively. *Past Productivity*$_{i;t,s}$ is the productivity of Author *i* during the period of Years $t-4$ to *t*. *Avg. co-authors' past productivity*$_{i;t,s}$ is the average of past productivity of all the co-authors of Author *i* during the period of Years $t-4$ to *t*. *Degree of specialization*$_{i;t,s}$ is the Herfindahl Index of paper topics for papers published during the period of Years $t-4$ to *t*. All continuous variables are winsorized at the top and bottom 1% level. Year, career seniority and Ph.D granting university fixed effects are included and standard errors are adjusted for author level clustering. *, **, *** indicate significance at the 0.10, 0.05 and 0.01 levels, respectively.

ness is 4.186 ($t = 13.65$). Authors that lie in the critical position of connecting authors have access to more information and produce more in the future.

For control variables, the coefficients on *Past productivity* ($q^{p}_{i;t,s}$) are positive and significant in all three columns. Consistent with our expectation, authors who are productive in the past tend to be productive in the future. The coefficients on the *Avg. co-authors' past productivity* ($\bar{q}_{1i;t,s}$) are positive and significant in all three columns. Working with productive co-authors has a positive impact on an author's future productivity. The coefficients on *Degree of specialization* ($H_{i;t,s}$) are negative and significant in all three columns. Therefore, a high concentration in a few topics has a negative impact on future productivity.

6.3. Author network centrality and topic innovation

Here, we examine how centrality affects an author's innovation in research topics. Assume that Author *i* publishes Paper *p*. We compare Paper *p* with Paper *q* published before Paper *p*. Paper *p* covers *m* LDA topics, among which *n* topics are covered in Paper *q*. Then, Paper *p* explores ($m - n$) new topics relative to Paper *q* and we define *Percentage of New Topics*$_{i,p,q} = \frac{m-n}{m}$ as the percentage of new topics covered in Paper *p* relative to Paper *q*. Note that for the same Paper *p*, *Percentage of New Topics*$_{i,p,q}$ will change when a different Paper *q* is used as a comparison.

For each paper published in Year *t*, we compare it with the most recent papers published before Year *t*.[11] We then examine the impact of Author *i's* network centrality on $Percentage\ \ of\ \ New\ \ Topics_{i,p,q}$ using the following model:

$$Percentage\ \ of\ \ New\ \ Topics_{i,p,q,t+1} = \alpha_0 + \alpha_1 Centrality + \alpha_2 Percentage\ \ of\ \ New\ \ Co_{author\,i,p,q} + \alpha_3 New\ \ Co\_author's\ \ New\ \ Topics_{i,p,q} + \alpha_4 Gap_{i,p,q} + D'_{i,t}\omega + University_i + \mu_t + \epsilon_{i;t,s} \quad (11)$$

where $Percentage\ \ of\ \ New\ \ Topics_{i,p,q,t+1}$ is $Percentage\ \ of\ \ New\ \ Topics_{i,p,q}$ for Author *i's* Paper *p* published in Year $t+1$, compared with the most recent papers published before Year $t+1$.[12] We regress $Percentage\ \ of\ \ New\ \ Topics_{i,p,q,t+1}$ on Author *i*'s centrality measures based on publications from Years *t* – *s* + 1 to *t*.

For each *p*-*q* pair, we calculate the percentage of new co-authors in Paper *p* relative to Paper *q*:

$$Percentage\ \ of\ \ New\ \ Co\_author_{i,p,q} = \frac{Number\ \ of\ \ co-authors\ \ in\ \ paper\ \ p\ \ but\ \ not\ \ in\ \ paper\ \ q}{Number\ \ of\ \ total\ \ authors\ \ in\ \ paper\ \ p} \quad (12)$$

We expect that more new co-authors in Paper *p* relative to Paper *q* will lead to new topics being explored in Paper *p* relative to Paper *q*.

We further control for $New\ \ Co\_author's\ \ New\ \ Topics_{i,p,q}$, which is the number of new topics new co-authors in Paper *p* have done before. We define $New\ \ Co\_author's\ \ New\ \ Topics_{i,p,q}$ as follows. If Paper *p* has Authors *i*, *a*, *b*, *c* and Paper *q* has Authors *i*, *a* and *d*. *b* and *c* are defined as new co-authors in Paper *p* relative to Paper *q*. We count the number of topics that Authors *b* and *c* have done but Author *i* has not done before Paper *p* is published and define the total number of such topics as $New\ \ Co\_author's\ \ New\ \ Topics_{i,p,q}$. We expect that new topics that new co-authors have done will lead to new topics being explored in Paper *p* relative to Paper *q*.

As the time gap between Paper *p* and other most recent papers are different for different Author *i*, we control for $\mathrm{Gap}_{\mathrm{i,p,q}}$, which is the time gap between Paper *p* and Paper *q*. We also control for year, career seniority, μ_t, and Ph.D granting university, $University_i$, fixed effects.

We estimate Model (11) and report the results in Table 9. We calculate centrality measures based on co-authorship network from Years *t* – 4 to *t* (five-year window). The coefficients on centrality measures are all positive and significant in Table 9 (0.003, $t = 2.75$ for *Degree,* 14.503, $t = 2.24$ for *Closeness* and 0.643, $t = 2.40$ for *Betweenness*). These results suggest that the higher is an author's network centrality, the more new topics he will explore in future publications.

The effects on *Percentage of New Co-authors* and *New Co-author's New Topics* are also consistent with our expectation. The coefficients on *Percentage of New Co-authors* and *New Co-author's New Topics* are significantly positive in all three columns. These results suggest that more new co-authors and more new topics these new co-authors have done will lead to more new topics being explored by an author.

6.4. Endogeneity

Our centrality measures are calculated based on the co-authorship network formed via an author's past publications. If both an author's research output and network centrality are affected by common omitted variables, the association between author network centrality and research output can be biased. Reverse causality is another concern. Several theoretical papers examine industrial firms' incentives in forming alliance with other firms in innovation activities (Cowan et al., 2007; Baum et al., 2010). Similarly, co-authorship links do not form randomly in our setting. Productive authors can attract more co-authors and hence gain high centrality in the network. We address endogeneity through an instrument variable (IV) approach. For *Degree*

[11] For example, if the most recent year before Year *t* an author has publications is Year *t* – *l*, then we compare with all papers published in Year *t* – *l* and calculate $Percentage\ \ of\ \ New\ \ Topics_{i,p,q}$ for each comparison.

[12] Assume that Author *i* has 2 publications in Year *t* + 1, 0 publication in Year *t*, 0 publication in Year *t* – 1, and 3 publications in Year *t* – 2, then he will have 2×3 $Percentage\ \ of\ \ New\ \ Co_{authors\,i,p,q,t+1}$ observations in the regression.

Table 9
Network centrality and topic innovation.

	(1) *Percentage of New Topics*	(2) *Percentage of New Topics*	(3) *Percentage of New Topics*
Degree	0.003 (2.75)***		
Closeness		14.503 (2.24)**	
Betweenness			0.643 (2.40)**
Percentage of New Co-authors	0.095 (8.66)***	0.097 (8.77)***	0.096 (8.69)***
New Co-author's New Topics	0.001 (2.68)***	0.001 (2.52)**	0.001 (2.55)**
Gap	−0.008 (−3.32)***	−0.009 (−4.01)***	−0.009 (−3.93)***
Year fixed effects	Yes	Yes	Yes
Career seniority fixed effects	Yes	Yes	Yes
Granting university fixed effects	Yes	Yes	Yes
N	8895	8895	8895
Adj. R-sq	0.054	0.054	0.054

This table presents results on the effect of author network centrality on the topic innovation. We estimate the following model:

$$\begin{aligned}\textit{Percentage of New Topics}_{i,p,q,t+1} &= \beta_0 + \beta_1\textit{Centrality}_{i;t,s} + \beta_2\textit{Percentage of New Co-author}_{i,p,q}\\ &+\beta_3\textit{New Co\_author's New Topics}_{i,p,q} + \beta_4\textit{Gap}_{i,p,q} + \textit{Year Fixed Effects}\\ &+\textit{Career seniority fixed effects} + \textit{Ph.D Granting university fixed effects},\end{aligned}$$

where *Percentage of New Topics*$_{i,p,q,t+1}$ is percentage of new topics for Author *i's* Paper *p* published in Year *t* relative to Author *i's* most recent Paper *q*. *Centrality*$_{i;t,s}$ represents the centrality measures for Author *i* during the period of Years $t-4$ to *t*. We use *Degree, Closeness* and *Betweenness* to measure *Centrality* respectively. *Percentage of New Co-author*$_{i,p,q}$ is the percentage of new authors for Author *i's* Paper *p* relative to the author's most recent Paper *q*. *New Co_author's New Topics*$_{i,p,q}$ is the number of topics the new co-authors in Paper *p* relative to Paper *q* has done while Author *i* has never done before Year *t*. *Gap*$_{i,p,q}$ measures the time gap between the publication year of Paper *p* and Paper *q*. All continuous variables are winsorized at the top and bottom 1% level. Year, career seniority and Ph.D granting university fixed effects are included and standard errors are adjusted for author level clustering. *, **, *** indicate significance at the 0.10, 0.05 and 0.01 levels, respectively.

and *Closeness*, we use the number of authors that can be indirectly connected to an author through three steps as an instrumental variable. We define *Order 3* as the number of three-step connected authors. *Order 3* serves as a proxy for a pool of potential co-authors as they are close to the author. Therefore, *Order 3* should be positively associated with subsequent period *Degree*. In addition, when an author has more potential co-authors to work with, her distance to other co-authors in the subsequent period is also likely to be close. We expect *Order 3* to be also positively associated with *Closeness*. However, *Order 3* of the previous period should not directly affect research output as these authors do not directly collaborate with the author.

The instrumental variable estimation results when *Degree* is used as a centrality measure are reported in Panel A, Table 10. Columns (1) and (2) report results for research productivity. Column (1) presents the first stage results. Consistent with our expectation, the coefficient on *Order 3* is positive and significant (0.084, $t = 85.89$). Authors with a large number of three-step authors tend to have high *Degree*. Column (2) reports the second stage result for the instrumented *Degree* (*IV_Degree*). The coefficient1 on *IV_ Degree* is positive and significant (0.037, $t = 7.08$). Columns (3) and (4) report results for topic innovation. Similarly, the coefficient on *Order 3* in the first stage is positive and significant (0.094, $t = 45.42$). The coefficient on *IV_Degree* is positive and significant (0.003, $t = 2.12$).

The instrumental variable estimation results for *Closeness* are reported in Panel B, Table 10. Similarly, the coefficients on *Order 3* in the first stage estimations are both positive and significant (0.00002, $t = 72.37$ in Column (1) and 0.00001, $t = 29.04$ in Column (3)). The coefficients on *IV_Closeness* are both positive and significant for research productivity and topic innovation (172.659, $t = 7.31$ for *Productivity* in Column (2) and 26.453, $t = 2.12$ for *Percentage of New Topics* in Column (4)).

Table 10
Centrality and research output: Instrumental variable approach.

Panel A: *Degree* and Research Output

	Productivity		Topic Innovation	
	First Stage	Second Stage	First Stage	Second Stage
	(1)	(2)	(3)	(4)
Variables	*Degree*	*Productivity*	*Degree*	*Percentage of New Topics*
IV_Degree	–	0.037		0.003
	–	(7.08)***		(2.12)**
Order 3	0.084		0.094	
	(85.89)***		(45.42)***	
Past productivity $(q^{p}_{i;t,s})$	1.151	0.55		
	(37.01)***	(33.04)***		
Avg. co-authors' past productivity $(\bar{q}_{1i;t,s})$	−0.147	0.046		
	(−11.02)***	(7.76)***		
Degree of specialization $(H_{i;t,s})$	−1.745	−0.09		
	(−29.52)***	(−2.85)***		
Percentage of New Co-authors			0.709	0.095
			(7.82)***	(8.97)***
New Co-author's New Topics			−0.011	0.001
			(−5.59)***	(2.99)***
Gap			−0.513	−0.007
			(–32.33)***	(−3.07)***
Year fixed effects	Yes	Yes	Yes	Yes
Career seniority fixed effects	Yes	Yes	Yes	Yes
PhD Granting university fixed effects	Yes	Yes	Yes	Yes
N	23,216	23,216	8895	8895
Adj. R-sq		0.5005		0.054

Panel B: *Closeness* and Research Output

	Productivity		Topic Innovation	
	First Stage	Second Stage	First Stage	Second Stage
	(1)	(2)	(3)	(4)
Variables	*Closeness*	*Productivity*	*Closeness*	*Percentage of New Topics*
IV_Closeness	–	172.659		
	–	(7.31)***	–	26.453
Order 3	0.00002	–	–	(2.12)**
	(72.37)***	–	0.00001	–
Past productivity $(q^{p}_{i;t,s})$	0.00001	0.593		
	(1.24)	(40.80)***		
Avg. co-authors' past productivity $(\bar{q}_{1i;t,s})$	0.00016	0.012		
	(45.89)***	(1.45)		
Degree of specialization $(H_{i;t,s})$	−0.00014	−0.134		
	(−8.79)***	(−4.42)***		
Percentage of New Co-authors			0.00001	0.097
			(0.51)	(9.21)***
New Co-author's New Topics			0.00000	0.001
			(0.67)	(2.79)***
Gap			−0.00003	−0.008
			(−8.88)***	(−3.62)***
Year fixed effects	Yes	Yes	Yes	Yes
Career seniority fixed effects	Yes	Yes	Yes	Yes
PhD Granting university fixed effects	Yes	Yes	Yes	Yes
Constant	0.00173	−4.307	0.00052	0.738
	(15.06)***	(−99.67)***	(7.26)***	(23.23)***
N	23,216	23,216	8895	8895
Adj. R-sq	0.535			0.053

Panel C: *Betweenness* and Research Output

	Productivity		Topic Innovation	
	First Stage	Second Stage	First Stage	Second Stage
	(1)	(2)	(3)	(4)
Variables	*Betweenness*	*Productivity*	*Betweenness*	*Percentage of New Topics*
IV_Betweenness	–	32.470	–	0.452
	–	(8.08)***	–	(0.58)
Clustering Coefficient	−0.005	–	−0.010	–
	(−32.66)***	–	(−36.19)***	–
Past productivity ($q^{p}_{i;t,s}$)	0.003	0.439		
	(21.11)***	(15.39)***		
Avg. co-authors' past productivity ($\bar{q}_{1i;t,s}$)	0.001	0.029		
	(16.93)***	(3.41)***		
Degree of specialization ($H_{i;t,s}$)	0.000	−0.033		
	(1.32)	(−0.73)		
Percentage of New Co-authors			0.002	0.088
			(4.82)***	(7.33)***
New Co-author's New Topics			0.000	0.001
			(1.20)	(2.26)**
Gap			−0.001	−0.006
			(−8.29)***	(−2.37)**
Year fixed effects	Yes	Yes	Yes	Yes
Career seniority fixed effects	Yes	Yes	Yes	Yes
PhD Granting university fixed effects	Yes	Yes	Yes	Yes
Constant	−0.001	−0.316	0.016	0.779
	(−0.64)	(−1.75)*	(9.89)***	(27.76)***
N	14,680	14,680	6898	6898
Adj. R-sq		0.342		0.046

This presents results for the instrumental variable two-stage results for the association between three centrality measures and research output.

In Panel A, we present the results for the association between *Degree* and research output. We use *Order 3*, which is the number of three-step-connected authors as an instrumental variable for *Degree*. *IV_Degree* is the instrumented *Degree* from the first stage. In Panel B, we present the results for the association between *Closeness* and research output. We use *Order 3*, which is the number of three-step-connected authors as an instrumental variable for *Closeness*. *IV_Closeness* is the instrumented *Closeness* from the first stage. In Panel C, we present the results for the association between *Betweenness* and research output. We use *Cluster Coefficients* as an instrumental variable for *Betweenness*. *IV_Betweeness* is the instrumented *Betweenness* from the first stage.

In each panel, the results for productivity are reported in Columns (1) and (2) and the results for topic innovation are reported in Columns (3) and (4).

Productivity$_{i;t,s}$ is the 3-year future productivity of Author *i* in Year *t*. *Percentage of New Topics*$_{i,p,q,t+1}$ is the percentage of new topics for Author *i's* Paper *p* published in Year *t* relative to Author *i's* most recent Paper *q*. *Past Productivity*$_{i;t,s}$ is the productivity of Author *i* during the period of Years $t-4$ to *t*. *Avg. co-authors' past productivity*$_{i;t,s}$ is the average of past productivity of all the co-authors of Author *i* during the period of Years $t-4$ to *t*. *Degree of specialization*$_{i;t,s}$ is the Herfindahl Index of paper topics for papers published during the period of Years $t-4$ to *t*. *Percentage of New Co-author*$_{i,p,q}$ is the percentage of new authors for Author *i's* Paper *p* relative to the author's most recent Paper *q*. *New Co_author's New Topics*$_{i,p,q}$ is the number of topics the new co-authors in Paper *p* relative to Paper *q* has done while Author *i* has never done before Year *t*. *Gap*$_{i,p,q}$ measures the time gap between the publication year of Paper *p* and Paper *q*. All continuous variables are winsorized at the top and bottom 1% level. Year, career seniority and Ph.D granting university fixed effects are included. For productivity, estimate the two-stage Tobit model and standard errors are adjusted for two-stage estimation. For topic innovation, standard errors are adjusted for two stage estimation. *, **, *** indicate significance at the 0.10, 0.05 and 0.01 levels, respectively.

Finally, we use *Clustering Coefficient* as an instrumental variable for *Betweenness*. Note that *Betweenness* measures the extent to which the shortest link connecting any two authors in the network passes through a given author. On the one hand, *Clustering Coefficient* can affect *Betweenness. Clustering Coefficient* measures the extent to which an author's co-authors are also co-authors of each other. One extreme case is that, when an author's *Clustering Coefficient* equals zero, none of the author's co-authors are directly connected to each other. As a result, all the shortest ways of connecting any two of the author's co-authors will have to pass through the author. Another extreme scenario is that, when an author's *Clustering Coefficient* equals one, any two of the author's co-authors are directly connected. In this case, none of the shortest ways will go through the author. Therefore, *Clustering Coefficient* has a negative impact on an author's *Betweenness*. On the other hand, *Clustering Coefficient* is a result of an author's co-authors' interactions and is not under his control.

We report the instrumental variable estimation results in Panel C, Table 10. Due to the nature of *Clustering Coefficient*, only authors with *Degree* equal to or larger than two will have *Clustering Coefficient*. Our sample in this test is based on authors with two or more co-authors. Columns (1) and (2) report results for the impact of *Betweenness* on research productivity. In Column (1), consistent with our expectation, the coefficient on *Clustering Coefficient* is negative and significant (−0.005, $t = –32.66$). Column (2) reports the second stage result of the instrumented *Betweenness* (*IV_Betweenness*). The coefficient on *IV_Betweenness* is positive and significant (32.470, $t = 8.08$). Columns (3) and (4) report results on topic innovation. The coefficient on *Clustering Coefficient* for the first stage is negative and significant (−0.10, $t = −36.19$). The coefficients on *IV_Betweenness* is positive but insignificant (0.452, $t = 0.58$). Using the instrumental variable approach, the positive association between *Betweenness* and research productivity still holds.

In sum, using a more rigorous empirical design to deal with endogeneity does not change the general tone of our findings.

6.5. Author seniority and the impact of centrality

We next examine whether the impact of centrality differs with author seniority. We define an indicator variable, *Senior*, that equals one if an author's career seniority is higher than seventeen years, and zero otherwise. *Senior* is interacted with centrality measures and included in the regression.

Table 11
Author seniority and the impact of network centrality.

Panel A: Research productivity

Variables	(1)	(2)	(3)
	Productivity	*Productivity*	*Productivity*
Degree	0.019		
	(15.60)***		
Closeness		68.279	
		(22.05)***	
Betweenness			2.419
			(6.42)***
Senior_Degree	0.015		
	(9.82)***		
Senior_Closeness		12.169	
		(2.95)***	
Senior_Betweenness			4.441
			(8.35)***
Senior	−2.843	−2.876	−2.862
	(−421.17)***	(−381.87)***	(−453.09)***
Past productivity ($q^{p}_{i;t,s}$)	0.566	0.605	0.594
	(98.05)***	(102.64)***	(104.48)***
Avg. co-authors' past productivity ($\bar{q}_{1i;t,s}$)	0.050	0.038	0.053
	(18.55)***	(13.22)***	(19.66)***
Degree of specialization ($H_{i;t,s}$)	−0.116	−0.156	−0.167
	(−9.60)***	(−12.29)***	(−13.42)***
Year fixed effects	Yes	Yes	Yes
Career seniority fixed effects	Yes	Yes	Yes
PhD Granting university fixed effects	Yes	Yes	Yes
N	23,216	23,216	23,216
Pseudo R-sq	0.180	0.179	0.178

Panel B: Topic innovation

	(1) *Percentage of New Topics*	(2) *Percentage of New Topics*	(3) *Percentage of New Topics*
Degree	0.003		
	(2.16)**		
Closeness		13.472	
		(1.74)*	
Betweenness			0.439
			(1.30)
Senior_Degree	−0.000		
	(−0.23)		
Senior_Closeness		2.341	
		(0.23)	
Senior_Betweenness			0.455
			(0.96)
Senior	0.359	0.357	0.357
	(15.33)***	(14.10)***	(15.05)***
Percentage of New Co-authors	0.096	0.097	0.096
	(8.68)***	(8.76)***	(8.65)***
New Co-author's New Topics	0.001	0.001	0.001
	(2.66)***	(2.52)**	(2.62)***
Gap	−0.008	−0.009	−0.009
	(−3.32)***	(−4.00)***	(−3.93)***
Year fixed effects	Yes	Yes	Yes
Career seniority fixed effects	Yes	Yes	Yes
Granting university fixed effects	Yes	Yes	Yes
N	8895	8895	8895
Adj. R-sq	0.054	0.054	0.054

This table presents results on the effect of career seniority on the association between author network centrality and the research output. In panel A, we estimate the following model:

$$\begin{aligned}\textit{Productivity}_{i;t} &= \beta_0 + \beta_1\textit{Centrality}_{i;t,s} + \beta_2\textit{Past Productivity}_{i;t,s} + \beta_3\textit{Avg.coauthors' past productivity}_{i;t,s}\\ &\quad + \beta_4\textit{Degree of specialization}_{i;t,s} + \beta_5\textit{Senior}_{i;t} + \beta_6\textit{Senior\_Centrality}_{i;t,s} + \textit{Year Fixed Effects}\\ &\quad + \textit{Career seniority fixed effects} + \textit{Ph.D Granting university fixed effects}\end{aligned}$$

In panel B, we estimate the following model:

$$\begin{aligned}\textit{Percentage of New Topics}_{i,p,q,t+1} &= \beta_0 + \beta_1\textit{Centrality}_{i;t,s} + \beta_2\textit{Percentage of New Co-author}_{i,p,q}\\ &\quad + \beta_3\textit{New Co\_author's New Topics}_{i,p,q} + \beta_4\textit{Gap}_{i,p,q} + \beta_5\textit{Senior}_{i;t}\\ &\quad + \beta_6\textit{Senior\_Centrality}_{i;t,s} + \textit{Year Fixed Effects}\\ &\quad + \textit{Career seniority fixed effects} + \textit{Ph.D Granting university fixed effects}\end{aligned}$$

$\textit{Productivity}_{i;t,s}$ is the 3-year future productivity of Author *i* in Year *t*. $\textit{Percentage of New Topics}_{i,p,q,t+1}$ is the percentage of new topics for Author *i's* Paper *p* published in Year *t* relative to Author *i's* most recent Paper *q*. $\textit{Centrality}_{i;t,s}$ are centrality measures for Author *i* during the period of Years $t-4$ to *t*. We use *Degree, Closeness* and *Betweenness* to measure *Centrality* respectively. $\textit{Past Productivity}_{i;t,s}$ is the productivity of Author *i* during the period of Years $t-4$ to *t*. $\textit{Avg. co-authors' past productivity}_{i;t,s}$ is the average of past productivity of all the co-authors of Author *i* during the period of Years $t-4$ to *t*. $\textit{Degree of specialization}_{i;t,s}$ is the Herfindahl Index of paper topics for papers published during the period of Years $t-4$ to *t*. $\textit{Percentage of New Co-author}_{i,p,q}$ is the percentage of new authors for Author *i's* Paper *p* relative to the author's most recent Paper *q*. $\textit{New Co\_author's New Topics}_{i,p,q}$ is the number of topics the new co-authors in Paper *p* relative to Paper *q* has done while Author *i* has never done before Year *t*. $\textit{Gap}_{i,p,q}$ measures the time gap between the publication year of Paper *p* and Paper *q*. $\textit{Senior}_{i,t}$ is an indicator that equals one if Author *i*'s career seniority is higher than 17 years, and zero otherwise. $\textit{Senior\_Centrality}_{i;t,s}$ is the interaction of $\textit{Senior}_{i;t}$ and $\textit{Centrality}_{i;t,s}$. All continuous variables are winsorized at the top and bottom 1% level. Year, career seniority and Ph.D granting university fixed effects are included and standard errors are adjusted for author level clustering. *, **, *** indicate significance at the 0.10, 0.05 and 0.01 levels, respectively.

Results for productivity are reported in Panel A, Table 11. The coefficients on the centrality measures remain positive and significant. The coefficients *on Senior_Centrality* are all positive and significant (0.015, $t = 9.82$ for *Senior_Degree*, 12.169, $t = 2.95$ for *Senior_Closeness* and 4.441, $t = 8.35$ for *Senior_Betweenness*). Therefore, the positive association between centrality and productivity generally strengthens with more senior authors.

Table 12
Isolated authors and research output.

Panel A: Research productivity

Variables	*Productivity*
Isolated	−0.065
	(−16.30)***
Past productivity ($q^{p}_{i;t,s}$)	0.625
	(112.91)***
Avg. co-authors' past productivity ($\bar{q}_{1i;t,s}$)	0.047
	(17.36)***
Degree of specialization ($H_{i;t,s}$)	−0.151
	(−12.25)***
Year fixed effects	Yes
Career seniority fixed effects	Yes
PhD granting university fixed effects	Yes
N	23,216
Pseudo R-sq	0.178

Panel B: Topic innovation

	Percentage of New Topics
Isolated	−0.010
	(−0.96)
Percentage of New Co-authors	0.099
	(8.85)***
New Co-author's New Topics	0.001
	(2.55)**
Gap	−0.009
	(−4.28)***
Year fixed effects	Yes
Career seniority fixed effects	Yes
Granting university fixed effects	Yes
Constant	0.608
	(31.93)***
N	8895
Adj. R-sq	0.053

This table presents results of the association between isolated author and research output.
In panel A, we estimate the following model:

$$Productivity_{i;t} = \beta_0 + \beta_1 Isolated_{i;t,s} + \beta_2 Past\ Productivity_{i;t,s} + \beta_3 Avg.coauthors'\ past\ productivity_{i;t,s} + \beta_4 Degree\ of\ specialization_{i;t,s} + Year\ Fixed\ Effects + Career\ seniority\ fixed\ effects + Ph.D\ Granting\ university\ fixed\ effects$$

In panel B, we estimate the following model:

$$Percentage\ of\ New\ Topics_{i,p,q,t+1} = \beta_0 + \beta_1 Isolated_{i;t,s} + \beta_2 Percentage\ of\ New\ Co\text{-}author_{i,p,q} + \beta_3 New\ Co\_author's\ New\ Topics_{i,p,q} + \beta_4 Gap_{i,p,q} + Year\ Fixed\ Effects + Career\ seniority\ fixed\ effects + Ph.D\ Granting\ university\ fixed\ effects,$$

$Isolated_{i;t,s}$ is an indicator equals one if Author *i* is an isolated author during the period of Years $t-4$ to *t*, and zero otherwise. $Productivity_{i;t,s}$ is the 3-year future productivity of Author *i* in Year *t*. *Percentage of New Topics*$_{i,p,q,t+1}$ is the percentage of new topics for Author *i's* Paper *p* published in Year *t* relative to Author *i's* most recent Paper *q*. *Past Productivity*$_{i;t,s}$ is the productivity of Author *i* during the period of Years $t-4$ to *t*. *Avg. co-authors' past productivity*$_{i;t,s}$ is the average of past productivity of all the co-authors of Author *i* during the period of Years $t-4$ to *t*. *Degree of specialization*$_{i;t,s}$ is the Herfindahl Index of paper topics for papers published during the period of Years $t-4$ to *t*. *Percentage of New Co-author*$_{i,p,q}$ is the percentage of new authors for Author *i's* Paper *p* relative to the author's most recent Paper *q*. *New Co_author's New Topics*$_{i,p,q}$ is the number of topics the new co-authors in Paper *p* relative to Paper *q* has done while Author *i* has never done before Year *t*. $Gap_{i,p,q}$ measures the time gap between the publication year of Paper *p* and Paper *q*. All continuous variables are winsorized at the top and bottom 1% level. Year, career seniority and Ph.D granting university fixed effects are included and standard errors are adjusted for author level clustering. *, **, *** indicate significance at the 0.10, 0.05 and 0.01 levels, respectively.

We next examine how an author's seniority affects the association between the centrality and topic innovation. We report the results in Panel B, Table 11. The coefficients on *Degree* (0.003, $t = 2.16$) and *Closeness* (13.472, $t = 1.74$) are positive and significant but the coefficient on *Betweenness* is insignificant (0.439, $t = 1.30$). The coefficients on the interactions *Senior_Degree* (−0.000, $t = -0.23$), *Senior_Closenss* (2.341, $t = 0.23$) and *Senior_Betweenness* (0.455, $t = 0.96$) are all insignificant. Therefore, it appears that author seniority does not exert an impact on the positive association between centrality and topic innovation.

6.6. Isolated authors

Here, we examine the association between isolated authors and their research output. We are interested in whether isolated authors have different research output compared with connected authors as isolated authors may adopt different strategies or work styles compared with other researchers. We define an indicator variable *Isolate* that equals one if an author is isolated, and zero otherwise. We replace *Centrality* in Models (7) and (11) with *Isolate*. The regression results are reported in Table 12. Panel A presents results for *Productivity*. The coefficient on *Isolate* is negative and significant (−0.065, $t = -16.30$), suggesting that isolated authors have lower productivity. Panel B reports the results for topic innovation. The coefficient on *Isolate* is negative but insignificant (−0.10, $t = -0.96$). Therefore, while isolated authors have lower productivity, they do not lag in topic innovation.

6.7. Co-Authors' network centrality

Finally, we examine whether co-authors' network centrality has an impact on an author's research output. We are interested in the effect of the centrality difference between an author and their co-authors. We define three centrality difference measures. *Dif_Degree* is defined as the average co-authors' *Degree* minus an author's

Table 13
Co-author network centrality and research output.

Panel A: Research productivity

Variables	(1)	(2)	(3)
	Productivity	*Productivity*	*Productivity*
Degree	0.031		
	(26.15)***		
Closeness		70.703	
		(23.70)***	
Betweenness			5.863
			(17.28)***
Dif_Degree	0.008		
	(12.50)***		
Dif _Closeness		33.762	
		(11.62)***	
Dif _Betweenness			2.859
			(12.02)***
Past productivity ($q^{p}_{i;t,s}$)	0.578	0.612	0.598
	(101.32)***	(105.74)***	(106.65)***
Avg. co-authors' past productivity ($\bar{q}_{1i;t,s}$)	0.036	0.032	0.044
	(12.03)***	(11.28)***	(15.48)***
Degree of specialization ($H_{i;t,s}$)	−0.109	−0.142	−0.164
	(−9.00)***	(−11.21)***	(−13.43)***
Year fixed effects	Yes	Yes	Yes
Career seniority fixed effects	Yes	Yes	Yes
Granting university fixed effects	Yes	Yes	Yes
N	23,216	23,216	23,216
Pseudo R-sq	0.180	0.179	0.178

Panel B: Topic innovation

	(1) *Percentage of New Topics*	(2) *Percentage of New Topics*	(3) *Percentage of New Topics*
Degree	0.005 (3.22)***		
Closeness		14.124 (2.11)**	
Betweenness			1.517 (3.40)***
Dif_Degree	0.003 (2.16)**		
Dif_Closeness		1.843 (0.24)	
Dif_Betweenness			1.218 (2.45)**
Percentage of New Co-authors	0.097 (8.75)***	0.098 (8.80)***	0.097 (8.72)***
New Co-author's New Topics	0.001 (2.54)**	0.001 (2.52)**	0.001 (2.42)**
Gap	−0.008 (−3.32)***	−0.009 (−3.96)***	−0.009 (−3.97)***
Year fixed effects	Yes	Yes	Yes
Career seniority fixed effects	Yes	Yes	Yes
Granting university fixed effects	Yes	Yes	Yes
Constant	0.589 (29.40)***	0.596 (29.47)***	0.608 (32.10)***
N	8895	8895	8895
Adj. R-sq	0.054	0.054	0.054

This table presents results on the effect of co-author network centrality on the research output.

In panel A, we estimate the following model:

$$Productivity_{i;t} = \beta_0 + \beta_1 Centrality_{i;t,s} + \beta_2 Past\ Productivity_{i;t,s} + \beta_3 Avg.coauthors'\ past\ productivity_{i;t,s}$$
$$+\beta_4 Degree\ of\ specialization_{i;t,s} + \beta_5 Dif\_Centrality_{i;t,s} + Year\ Fixed\ Effects$$
$$+Career\ seniority\ fixed\ effects + Ph.D\ Granting\ university\ fixed\ effects$$

In panel B, we estimate the following model:

$$Percentage\ of\ New\ Topics_{i,p,q,t+1} = \beta_0 + \beta_1 Centrality_{i;t,s} + \beta_2 Percentage\ of\ New\ Co\text{-}author_{i,p,q}$$
$$+\beta_3 New\ Co\_author's\ New\ Topics_{i,p,q} + \beta_4 Gap_{i,p,q} + \beta_5 Dif\_Centrality_{i;t,s}$$
$$+Year\ Fixed\ Effects + Career\ seniority\ fixed\ effects + Ph.D\ Granting\ university\ fixed\ effects$$

In the two models, $Productivity_{i;t}$ is the 3-year future productivity of Author *i* in Year *t*. $Percentage\ of\ New\ Topics_{i,p,q,t+1}$is percentage of new topics for Author *i's* Paper *p* published in Year *t* relative to Author *i's* most recent Paper *q*. *Centralityi;t,s* represents the centrality measures for Author *i* during the period of Years *t* – 4 to *t*. We use *Degree, Closeness* and *Betweenness* to measure *Centrality* respectively. $Past\ Productivity_{i;t,s}$ is the productivity of Author *i* during the period of Years *t* – 4 to *t*. $Avg.\ co\text{-}authors'\ past\ productivity_{i;t,s}$ is the average of past productivity of all the co-authors of Author *i* during the period of Years *t* – 4 to *t*. $Degree\ of\ specialization_{i;t,s}$ is the Herfindahl Index of paper topics for papers published during the period of Years *t* – 4 to *t*. $Percentage\ of\ New\ Co\text{-}author_{i,p,q}$ is the percentage of new authors for Author *i's* Paper *p* relative to the author's most recent Paper *q*. $New\ Co\_author's\ New\ Topics_{i,p,q}$ is the number of topics the new co-authors in Paper *p* relative to Paper *q* has done while Author *i* has never done before Year *t*. $Gap_{i,p,q}$ measures the time gap between the publication year of Paper *p* and Paper *q*. $Dif\_Centrality_{i;t,s}$ is Author *i*'s co-authors' average *Centrality measures* during the period of Years *t* – 4 to *t*. All continuous variables are winsorized at the top and bottom 1% level. Year, career seniority and PhD granting university fixed effects are included and standard errors in parentheses are adjusted for author level clustering. *, **, *** indicate significance at the 0.10, 0.05 and 0.01 levels, respectively.

Degree. *Dif_Closeness* is defined as the average co-authors' *Closeness* minus an author's *Closeness*. *Dif_Betweenness* is the average co-authors' *Betweenness* minus an author's *Betweenness*. If an author has no co-authors, *Dif_Degree*, *Dif_Closeness* and *Dif_Betweenness* are set to zero.

We include these variables in the regression analyses and results are reported in Table 13. In Panel A, Table 13, we report the results for productivity. The coefficients on centrality measures are all positive and significant. In addition, the coefficients on *Dif_Degree*, *Dif_Closeness* and *Dif_Betweenness* are all positive and significant (0.008, $t = 12.50$ for *Dif_Degree*, 33.762, $t = 11.62$ for *Dif_Closeness* and 2.859, $t = 12.02$ for *Dif_Betweenness*). Thus, co-authors' centrality, relative to an author's own centrality, also have a positive impact on research productivity.

Panel B of Table 13 reports the results for topic innovation. The coefficients on the three centrality measures are all positive and significant. The coefficients on *Dif_Degree*, *Dif_Closeness*, and *Dif_Betweenness* are positive, but they are significant for *Dif_Degree* and *Dif_Betweenness* only (0.003, $t = 2.16$ for *Dif_Degree*, 1.843, $t = 0.24$ for *Dif_Closeness*, and 1.218, $t = 2.45$ for *Dif_Betweenness*). Overall, these results suggest that co-authors' centrality relative to an author's own centrality also generally positively affect topic innovation.

7. Conclusion

We establish the network centrality generated through research collaboration and examine the impact of centrality on individual researchers' output. Using papers published in the five top accounting journals (*JAR*, *TAR*, *JAE*, *CAR*, and *RAS*) from 1980 to 2016, we explore characteristics of the co-authorship network, research topic development and the impact of co-authorship network centrality on research output. We show that the co-authorship network in the accounting field meets the four features of the "small world" property. Specifically, each author does not have many co-authors compared with the number of authors in the network. There exists a giant component in the network within which any two authors can be linked, directly or indirectly. Further, it usually takes only a few steps to connect two authors in the network. Finally, the co-authorship in the network is highly overlapping. We further identify a group of authors that have very high co-authorship, and find that for these authors, their co-authors are less likely to work with each other. Therefore, the most connected authors play an important role in sustaining the network.

We use the LDA machine learning modeling to automatically label a research paper with multiple topics. Based on the LDA topic labels, we find that the number of topics covered each year relative to the number of published papers has decreased in recent years. In addition, the top-10 topics in each year are sticky. We also find that the overlaps of topics in the five journals are low, suggesting that these journals have their own topic specializations and/or tastes.

Finally, we examine the association between centrality in the co-authorship network and research output. We find that high centrality is associated with high future research productivity and topic innovation. We use an instrumental variable approach to address endogeneity associated with our centrality measures and find similar results. We execute several further analyses on the association between network features and research output. We find that author seniority enhances the positive impact of centrality on research productivity but not topic innovation. Isolated authors exhibit lower research productivity, but their do not lag in topic innovation. Finally, centrality of an author's co-authors also has an incrementally positively impact on his research output. Overall, we conclude that network centrality positively influence research output.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Ahuja, M.K., Galletta, D.F., Carley, K.M., 2003. Individual centrality and performance in virtual R&D groups: an empirical study. Manage. Sci. 49 (1), 21–38.

Akbas, F., Meschke, F., Wintoki, M.B., 2016. Director networks and informed traders. J. Account. Econom. 62, 1–23.

Allen, T.J., 1971. Communication networks in R&D laboratories. R&D Manage. 1, 14–21.

Allen, T.J., 1977. Managing the flow of technology. MIT Press, Cambridge.

Allen, J., James, A.D., Gamlen, P., 2007. Formal versus informal knowledge networks in R&D: A case study using social network analysis. R&D Manage. 37, 179–196.

Baum, J.A.C., Cowan, R., Jonard, N., 2010. Network-independent partner selection and the evolution of innovation networks. Manage. Sci. 56 (11), 2094–2110.

Becker, W., Dietz, J., 2004. R&D cooperation and innovation activities of firms- evidence for the german manufacturing industry. Res. Policy 33, 209–223.

Blei, D., Latterty, J., 2009. Topic models. In: Srivastava, A., Sahami, M. (Eds.), Text Mining: Classification, Clustering and Applications. Chapman & Hall/CRC Press.

Blei, D.M., Ng, A.Y., Jordan, M.I., 2003. Latent dirichlet allocation. J. Mach. Learn. Res. 3, 993–1022.

Bonner, S.E., Hesford, J.W., van der Stede, W.A., Young, S.M., 2012. The social structure of communication in major accounting research journals. Contemp. Account. Res. 29 (3), 869–909.

Chullun, T., Prevost, A., Puthenpurackal, J., 2014. Board ties and the cost of corporate debt. Financ. Manage. 43 (3), 533–568.

Cowan, R., Jonard, N., Zimmermann, J., 2007. Bilateral collaboration and the emergence of innovation networks. Manage. Sci. 53 (7), 1051–1067.

de Faria, P., Lima, F., Santos, R., 2010. Cooperation in innovation activities: the importance of partners. Res. Policy 39, 1082–1092.

Ductor, L., 2015. Does co-authorship lead to higher academic productivity? Oxford Bull. Econ. Stat. 77 (3), 0305–9049.

Ductor, L., Fafchamps, M., Goyal, S., van der Leij, M.J., 2014. Social networks and research output. Rev. Econom. Statist. 96 (5), 936–948.

El-Khatib, R., Fogel, K., Jandik, T., 2015. CEO network centrality and merger performance. J. Financ. Econ. 116, 349–382.

Fafchamps, M., van der Leij, M.J., Goyal, S., 2010. Matching and network effects. J. Eur. Econom. Assoc. 8 (1), 203–231.

Faleye, O., Kovacs, T., Venkateswaran, A., 2014. Do better-connected CEOs innovate more? J. Financ. Quant. Anal. 49, 1201–1225.

Freeman, L.C., 1977. A set of measures of centrality based on betweenness, 40 (1), 35–41.

Glover, S.M., Prawitt, D.F., Wood, D.A., 2006. Publication records of faculty promoted at the top 75 accounting research programs. Issues Account. Educ. 21 (3), 195–218.

Glover, S.M., Prawitt, D.F., Summers, S.L., Wood, D.A., 2012. Publication benchmarking data based on faculty promoted at the Top 75 U.S. accounting research institutions. Issues Account. Educ. 27 (3), 647–670.

Goldenberg, J., Libai, B., Muller, E., Stremersch, S., 2010. Database submission: the evolving social network of marketing scholars. Market. Sci. 29 (3), 561–567.

Goyal, S., van der Leij, M.J., Moraga-Gonzalez, J.L., 2006. Economics: an emerging small world. J. Polit. Econ. 14 (2), 403–412.

Griffiths, T.L., Steyvers, M., 2004. Finding scientific topics. Proc. Natl. Acad. Sci. United States America 101 (1), 1915–2016.

Grun, B., Hornik, K., 2011. Topicmodels: An R Package for Fitting Topic Models. Available at: https://cran.r-project.org/web/packages/topicmodels/vignettes/topicmodels.pdf.

Hall, D., Jurafsky, D., Manning, C.D., 2008. Studying the history of ideas using topic models. In: EMNLP '08 Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 363–371.

Hasselback, J., Reinstein, A., Schwan, E.S., 2000. Benchmarks for evaluating the research productivity of accounting quality. J. Account. Educ. 18 (2), 79–97.

Hollis, A., 2001. Co-authorship and the output of academic economists. Labor Econom. 8 (4), 503–530.

Larcker, D., So, E.C., Wang, C.C.Y., 2013. Boardroom centrality and firm performance. J. Account. Econom. 55 (2–3), 225–250.

Lohmann, C., Eulerich, M., 2017. Publication trends and the network of publication institutions in accounting: data on the accounting review, 1926–2014. Account. History Rev. 27 (1), 1–25.

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

Oler, D.K., Oler, M.J., Skousen, C.J., 2010. Characterizing accounting research. Account. Horiz. 24 (4), 635–670.

Sabidussi, G., 1966. The centrality index of a graph. Psychometrika 31 (4), 581–603.

Watts, D.J., Strogatz, S.H., 1988. Collective dynamics of "Small-world" networks. Lett. Nat. 393, 440–442.