



**WP/20/144**

# **IMF Working Paper**

---

## **Network Effects and Research Collaborations**

by Dennis Essers, Francesco Grigoli, Evgenia Pugacheva

I N T E R N A T I O N A L M O N E T A R Y F U N D

# Network Effects and Research Collaborations\*

Dennis Essers<sup>†</sup>      Francesco Grigoli<sup>‡</sup>      Evgenia Pugacheva<sup>§</sup>

## Abstract

We study the determinants of new and repeated research collaborations, drawing on the co-authorship network of the International Monetary Fund (IMF)’s Working Papers series. Being an outlet where authors express their views on topics of interest, and given that IMF staff is not subject to the “publish-or-perish” conditions of the academia, the IMF Working Papers series constitutes an appropriate testing ground to examine the endogenous nature of co-authorship formation. We show that the co-authorship network is characterized by many authors with few direct co-authors, yet indirectly connected to each other through short co-authorship chains. We find that a shorter distance in the co-authorship network is key for starting research collaborations. Also, higher research productivity, being employed in the same department, and having citizenship of the same region help to start and repeat collaborations. Furthermore, authors with different co-authorship network sizes are more likely to collaborate, possibly reflecting synergies between senior and junior staff members.

**Keywords:** Research collaboration, co-authorship, networks, research publications, IMF Working Papers.

**JEL Codes:** D85, O31.

---

\*The views expressed in this working paper are those of the authors and do not necessarily represent those of the IMF, its Executive Board, its management, the National Bank of Belgium, or the Eurosystem. Working papers describe research in progress by the authors and are published to elicit comments and to encourage debate. We thank, without implicating, Roberto Guimaraes-Filho and his team for compiling the employment and demographic data used in this paper.

<sup>†</sup>National Bank of Belgium, Economics and Research Department, [dennis.essers@nbb.be](mailto:dennis.essers@nbb.be).

<sup>‡</sup>International Monetary Fund, Research Department, [fgrigoli@imf.org](mailto:fgrigoli@imf.org).

<sup>§</sup>International Monetary Fund, Research Department, [epugacheva@imf.org](mailto:epugacheva@imf.org).

# Contents

|  | Page      |
|--|-----------|
| <b>1 Introduction</b> . . . . .                      | <b>4</b>  |
| <b>2 IMF Research Under the Microscope</b> . . . . . | <b>7</b>  |
| 2.1 Data . . . . .                                   | 7         |
| 2.2 Stylized Facts . . . . .                         | 8         |
| 2.2.1 Publications . . . . .                         | 8         |
| 2.2.2 Authors . . . . .                              | 9         |
| 2.2.3 Author Pairs . . . . .                         | 13        |
| 2.2.4 Publication Length and Citations . . . . .     | 13        |
| 2.2.5 Research Topics . . . . .                      | 15        |
| <b>3 Co-Authorship Network</b> . . . . .             | <b>19</b> |
| 3.1 Author-Level Network . . . . .                   | 19        |
| 3.2 Institution-Level Network . . . . .              | 24        |
| 3.3 Evolution of the Network . . . . .               | 29        |
| <b>4 Determinants of Collaborations</b> . . . . .    | <b>34</b> |
| 4.1 Empirical Strategy . . . . .                     | 34        |
| 4.2 Results . . . . .                                | 36        |
| 4.2.1 Robustness . . . . .                           | 40        |
| <b>5 Conclusions</b> . . . . .                       | <b>42</b> |
| <b>References</b> . . . . .                          | <b>44</b> |

## List of Figures

|  | <b>Page</b> |
|--|-------------|
| 1 Lifetime of a Pair of Authors . . . . .  | 8           |
| 2 Collaboration Characteristics . . . . .  | 10          |
| 3 Author Characteristics . . . . .   | 12          |
| 4 Pair Characteristics . . . . .   | 14          |
| 5 Publication Lengths and Citations . . . . .  | 16          |
| 6 Publication Topics . . . . .   | 18          |
| 7 Full Author-Level Network, 1990–2017 . . . . .   | 20          |
| 8 Author Degrees vs. Clustering Coefficients . . . . .   | 22          |
| 9 Core Author-Level Network, 1990–2017 . . . . .   | 23          |
| 10 Full Institution-Level Network, 1997–2017 . . . . .   | 25          |
| 11 Core Institution-Level Network, 1997–2017 . . . . .   | 26          |
| 12 Full Network of IMF Departments and Other Institutions, 1999–2017 . . . . .                       | 27          |
| 13 Heat Map of Collaborations between IMF Departments and Other Institutions,<br>1999–2017 . . . . . | 28          |
| 14 Evolution of 10-year Cumulative Author-Level Network . . . . .                                    | 30          |
| 15 Distributions of Network Statistics (Density units) . . . . .                                     | 31          |
| 16 Evolution of 10-year Cumulative Core Author-Level Network . . . . .                               | 32          |
| 17 Clustering of Core Author-Level Network, 2010–2017 . . . . .                                      | 33          |
| 18 ROC Curves . . . . .  | 40          |

## List of Tables

|  | <b>Page</b> |
|--|-------------|
| 1 Network Statistics . . . . .   | 21          |
| 2 Determinants of First Collaborations . . . . .                           | 38          |
| 3 Determinants of Subsequent Collaborations . . . . .                      | 39          |
| 4 Determinants of First and Subsequent Collaborations—Robustness . . . . . | 41          |

# 1 Introduction

Knowledge is increasingly produced in teams, and, as in other social sciences, research collaborations in economics have steadily gained importance in recent decades (Wuchty et al., 2007; Henriksen, 2016).<sup>1,2</sup> Teaming up allows researchers to exploit synergies and economies of scale by pooling ideas, skills, time, and funds. While there is no consensus on the effect of co-authorship relations on the amount and quality of research output, most recent studies that control for the endogenous nature of co-authorship formation find a positive relation with overall and authors' individual productivity (Ductor et al., 2014; Ductor, 2015; Sommer and Wohlrabe, 2017).<sup>3</sup> Moreover, through collaborations, knowledge spills over directly to co-authors (Azoulay et al., 2010; Borjas and Doran, 2015), as well as indirectly to the broader network of collaborating researchers (Hsieh et al., 2018). But how do such collaboration networks look like in practice? And what are the determinants of starting a research collaboration? Are these the same factors that lead researchers to continue working together?

To address these questions, we construct the co-authorship network of the near universe of International Monetary Fund (IMF) Working Papers published over almost three decades (1990–2017), and combine it with authors' employment history and demographic characteristics. The IMF Working Papers series is one of the main research outlets of the institution and is widely read in central banks, by government officials, academics, and researchers at think tanks.<sup>4</sup>

Compared to other IMF research publications, the working papers express the authors' views rather than the institutional ones, can cover the geographical area and topic of interest of the authors, and do not restrain authorship to people within the same division or department.<sup>5</sup> Another important aspect is that IMF staff—the main authors of the IMF Working Papers—does not face the same “publish-or-perish” incentives of the academia. Hence, collaborations likely reflect personal preferences (e.g., researchers sharing a similar background) or practical considerations (e.g., researchers working in the same department), rather than strategic ones (e.g., co-authorship with internationally renowned researchers that can help get the papers into top-tier journals). Moreover, the IMF Working Papers series constitutes a potentially important channel for collaboration with outside researchers from academia, country authorities, other international organizations, and

---

<sup>1</sup>Kuld and O'Hagan (2018), for example, show that the share of multi-authored papers published in a large sample of economics journals rose from around 50 percent in 1996 to almost 75 percent in 2014, mostly due to a pickup in papers with three or more co-authors (see also Card and DellaVigna, 2013; Hamermesh, 2013; Rath and Wohlrabe, 2016).

<sup>2</sup>Throughout the paper we will use co-authorship as a proxy for research collaboration. As Adams (2012, p. 335) notes, since few researchers would casually surrender credit for their publications, one can assume that sharing authorship generally reflects a tangible engagement.

<sup>3</sup>Earlier studies that found no significant or even a negative effect include Hollis (2001) and Medoff (2003).

<sup>4</sup>Officials report using IMF Working Papers frequently as a reference in policy discussions. At the IMF itself they are widely read by staff and serve as “a vehicle to disseminate emerging ideas within the institution, to share new types of analysis and new ways of looking at country policies” (IEO, 2011, p. 12). A bench-marking study by Aizenman et al. (2011) finds that, over the years 1999–2009, IMF Working Papers were cited less often than those of the Federal Reserve, but more than those of other comparators such as the World Bank, the Organisation for Economic Co-operation and Development, and the Inter-American Development Bank.

<sup>5</sup>Other IMF research publications include the flagships (*World Economic Outlook*, *Global Financial Stability Report*, *Regional Economic Outlooks*, *Fiscal Monitor*, and *External Sector Report*), as well as *Staff Discussion Notes*, *Departmental Papers*, *G-20 Surveillance Notes*, and *Spillover Notes*, among others (IEO, 2011).

think tanks. Finally, by focusing on working papers rather than on peer-reviewed journal articles, the problem of publication lags distorting our findings is much less severe.<sup>6</sup> In light of these considerations, we argue that the network of the IMF Working Papers co-authorship provides an appropriate testing ground to examine the questions at hand, since it should reflect the endogenous nature of how collaborations between researchers are initiated and maintained.

In this paper, we first document an extensive set of stylized facts on the number, length, co-authorship, topics, and citation counts of IMF Working Papers; the employment and demographic characteristics of their IMF authors; and how all this has evolved over time. We confirm that—similar to the trends observed in economics journals—the number of co-authors per paper has steadily increased over time. The distribution of the number of papers written by any given author, however, is highly right-skewed, with the large majority of authors having published only one paper and very few producing more than 20. Unsurprisingly, the Research Department accounts for the most papers in absolute numbers and for the largest share of publishing economists by department; its papers also tend to get more citations. Moreover, IMF research authorship has become more diverse, reflecting changes in its workforce, i.e., women and people with non-US and non-European backgrounds increased their shares in total working paper output. In terms of topic coverage, unsurprisingly, most working papers deal with macroeconomics, international economics, financial economics, and public economics. However, in recent years, increased attention has been dedicated to less conventional or newer topics for the IMF, including climate change and gender.

Next, we investigate the broader co-authorship network of the IMF Working Papers. We study the network both at the level of the individual authors and at the level of the IMF departments and external institutions these authors are affiliated with. The latter allows us to shed light on the extent to which research published at the IMF is the result of collaborations within departments, across the IMF, and with the outside world. Unsurprisingly, IMF staff dominates the network in terms of the number of authors and working papers, followed at a clear distance by academics, central bankers, and staff from other international organizations. Authors affiliated with government agencies, private companies, and non-profits occupy a much less prominent role in the network. Also, IMF staff has been collaborating with researchers from many different organizations within the same institutional category, but at the internal margin these collaborations are very heterogeneous. The most recurrent links are with researchers from the World Bank and U.S. and U.K. universities, and, in second instance, from other European universities and advanced economy central banks. Among the institutions involved in at least ten papers, there are only a few from emerging market or developing economies and no government agencies or non-profits. Research collaborations within the same IMF department are more common than co-authorship across different departments or with external institutions.

Network statistics, including the size of the giant component and the average degree, distance, and clustering coefficient, confirm that the co-authorship network has not only become larger over time but also more integrated. Taken together, our findings indicate that the IMF Working Papers network exhibits “small world” properties. That is, the network consists of a large set of au-

---

<sup>6</sup>IMF Working Papers do undergo an internal review process but the time between when the paper is finalized and when it is published is typically much shorter than in the case of articles in peer-reviewed journals.

thors with typically few direct co-authors; the majority of authors are nonetheless in some way connected to each other, mostly through short chains of co-authors; and the degree of overlap in co-authorship is very high (Newman, 2001; Goyal et al., 2006). As Goyal et al. (2006) and Hsieh et al. (2018) show, similar patterns are also found in the network of economics journals and derive from a network architecture of interlinked “star” authors that act as connectors of different clustered parts of the network.

Finally, we empirically investigate how IMF Working Papers co-authorship relations are formed and maintained, which could be different from how the purely academic collaborations, on which the literature focuses, come about (see e.g., McDowell and Melvin, 1983; Sutter and Kocher, 2004; Liben-Nowell and Kleinberg, 2007; Frenken et al., 2009; Freeman and Huang, 2015; Colussi, 2018; Hsieh et al., 2018). Methodologically, our analysis of the determinants of research collaborations is closest to the work of Fafchamps et al. (2010), who study co-authorship in a large sample of peer-reviewed articles published during 1970–1999 in economics journals indexed in the EconLit database. In addition to the characteristics analyzed in Fafchamps et al. (2010), we look at the IMF authors’ gender, region of citizenship, and the department in which they are employed. This extra information allows us to test additional hypotheses and to control for possible confounding factors when making inference on the effect of pure network variables on the probability of collaborating.

We confirm the main finding of Fafchamps et al. (2010): above and beyond other pair and individual author characteristics, two researchers that are closer to each other in the existing co-authorship network are significantly more likely to collaborate, especially in the case of first-time collaborations. This result is in line with the interpretation that collaboration networks may transmit important, not directly observable information about authors (the so-called network effects) that helps to overcome matching frictions. Once this information is internalized through actual collaboration, the network effects strongly diminish. In addition, we find that more productive authors are more likely to start and continue collaborations. At the same time, researchers with an established network of co-authors seem to be less in search of new ones. Yet, authors with different co-author network sizes are more likely to start and maintain a collaboration, possibly reflecting collaborations between senior staff members (generally tasked with more managerial duties) and junior staff (who have more time for research). Finally, we find that being employed in the same department and having citizenship of the same region help to start and repeat collaborations, while greater overlap in research areas does not.

The paper is structured as follows. Section 2 explains the construction of our dataset and presents key stylised facts about the papers, their authors, pairs of co-authors, citations, and the topics covered. Section 3 visualizes and describes the co-authorship network at the author and institutional levels. Section 4 investigates the determinants of first and subsequent collaborations across authors. Section 5 concludes.

## 2 IMF Research Under the Microscope

### 2.1 Data

We collected information on 6,152 IMF Working Papers written between 1990 and 2017 by 3,918 distinct authors.<sup>7</sup> Based on the information obtained from the IMF website catalog and the text of the publications, for each working paper we compiled the list of authors, and, where available, the number of pages, author emails, and the *Journal of Economic Literature* (JEL) codes that authors used to classify publications by economic research area(s). JEL codes are available starting in 1997.

We complemented the information from the working papers with employee-level data for the years 1999 to 2017. This unique dataset includes annual information on the job title, year of birth (in five-year brackets), tenure, department, citizenship (by region), and gender of IMF staff. As differences in author name spelling were often present between working papers and employee-level data, we used fuzzy matching and performed extensive manual checks to merge the data.<sup>8</sup> We then matched job titles to grade levels to distinguish between support-level staff such as research assistants (grades A01–A10), economists (grades A11–A15), and managerial staff (grades B01–B05).<sup>9</sup> We calculated the authors’ age at the time the publication was released based on the first year of the five-year birth bracket. For time-invariant author characteristics we extended the data back prior to 1999.

For non-IMF authors of IMF Working Papers published between 1997 and 2017 we compiled information on their main affiliation based on author emails, where available, and manual search online in authors’ CVs, LinkedIn profiles, and on other websites.<sup>10</sup> We organized these external affiliations into six mutually exclusive categories: universities, other international organizations, central banks, government agencies, non-profit organizations (including think tanks and independent research institutes), and private companies (including consultancy firms and financial institutions).

Finally, we collected citation counts for the working papers from Google Scholar.<sup>11</sup> Relative to other sources of citations—such as Citations in Economics (CitEc, linked to the Research Papers in Economics platform or RePEc), Web of Science, or Scopus—Google Scholar has broader coverage of citations in books, dissertations, and in the non-scholarly “grey” literature (e.g., government reports), as well as in non-English language publications (Martín-Martín et al., 2018). Such broader coverage seems appropriate in light of IMF Working Papers’ wider and relatively in-

---

<sup>7</sup>See <https://www.imf.org/en/publications/search?series=IMF+Working+Papers>, accessed June 2020. We excluded working papers for which no authors were listed (44 papers in total, all from the 1990s). In 1990, our records undercount the total number of working papers published since not all of them have been indexed on the IMF website. For subsequent years, the number of non-indexed working papers is negligible. We then downloaded the PDF files that contain the text of the working papers. PDF files were not available for 822 publications, primarily in the years prior to 1997.

<sup>8</sup>Besides spelling mistakes, instances of where such differences occurred include authors with middle names, female authors using maiden or married names, and authors using shorter nicknames and alternative, Westernized names.

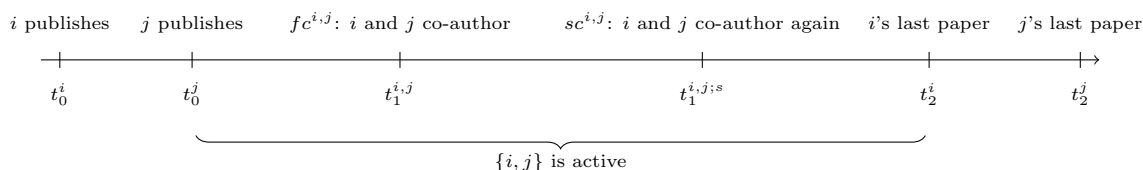
<sup>9</sup>See the “IMF Staff Salary Structure” available at <https://www.imf.org/external/pubs/ft/ar/2017/eng/assets/ar17-web-table-3-4.pdf>, accessed June 2020.

<sup>10</sup>We classified also IMF visiting and resident scholars as “non-IMF” authors and looked up their home universities.

<sup>11</sup>See <https://scholar.google.com>, accessed January 2019.



Figure 1: Lifetime of a Pair of Authors



Notes: The figure shows the lifetime of a pair of authors that co-author more than one paper.

ternational audience (IEO, 2011).<sup>12</sup> The citations we extracted are cumulative as of January 2019. To account for the fact that working papers were released at different times, we annualized the citation count by dividing it by the number of years since publication.

With the just-described data at hand, we define the universe of “active” potential co-authors in any given year  $t$  as  $S_t$ . This includes all those authors since their earliest year of publication,  $t_0$ , until their latest year of publication,  $t_2$ . A pair of authors  $\{i, j\} \in S_t$  is considered active from the year in which *both* authors are active,  $t_0^{i,j} = \max\{t_0^i, t_0^j\}$ , until the year in which *either* one of the two authors publishes his or her last paper,  $t_2^{i,j} = \min\{t_2^i, t_2^j\}$ . First collaborations are traced with a dummy variable,  $fc_t^{i,j}$ , that is defined only when the author pair is active, and that takes value one when authors  $i$  and  $j$  collaborate for the first time, and zero for the years before that. Subsequent collaborations are traced with another dummy variable,  $sc_t^{i,j}$ , which is defined only during the period in which the pair is active and after the first publication, and that takes value one in the year or years in which the pair publishes any other paper but the first, and zero otherwise. Figure 1 provides a visualization of the lifetime of a pair of authors which co-author more than one paper.

## 2.2 Stylized Facts

We now present some key stylized facts about IMF Working Papers, their authors, pairs of co-authors, the length and citation count of the papers, and the topics covered.

### 2.2.1 Publications

Figure 2 summarizes some features of the publications. As mentioned in the Section 2.1, between 1990 and 2017, 3,918 authors published a total of 6,152 IMF Working Papers. The number of publications per year increased steadily up to 2006 and stabilized at around 250 to 300 thereafter. Similarly, the number of authors increased, but at a faster rate and with no evidence of plateauing, reaching almost 600 in 2017 (panel 2a).

Consistent with the number of authors outpacing the number of publications, the number of authors per working paper increased over time (panel 2b), a trend also observed at top-tier economics journals and research publications in other disciplines (Card and DellaVigna, 2013; Hamermesh, 2013; Kuld and O’Hagan, 2018). In 2017, three fourths of the publications were conducted

<sup>12</sup>For similar reasons, Ravallion and Wagstaff (2012) use Google Scholar to collect citations for the World Bank’s research publications.

in collaboration with someone else. Yet, collaborations are not ubiquitous, as over their entire publishing career most authors worked with just one to three co-authors, though some researchers collaborated with over 100 distinct co-authors (panel 2c).

In any given year, about 80 percent of the authors wrote just one paper (panel 2d) and around one fourth of publications was produced by first-time authors (panel 2e). Moreover, the large majority of authors wrote only one paper over their publishing career at the IMF. For those who wrote more than one paper, the median number of papers was four, with very few authors producing more than 20 papers (panel 2f). The number of years an author remained active—excluding one-off papers—was six on average, with half of the authors being active for five years or less.

### 2.2.2 Authors

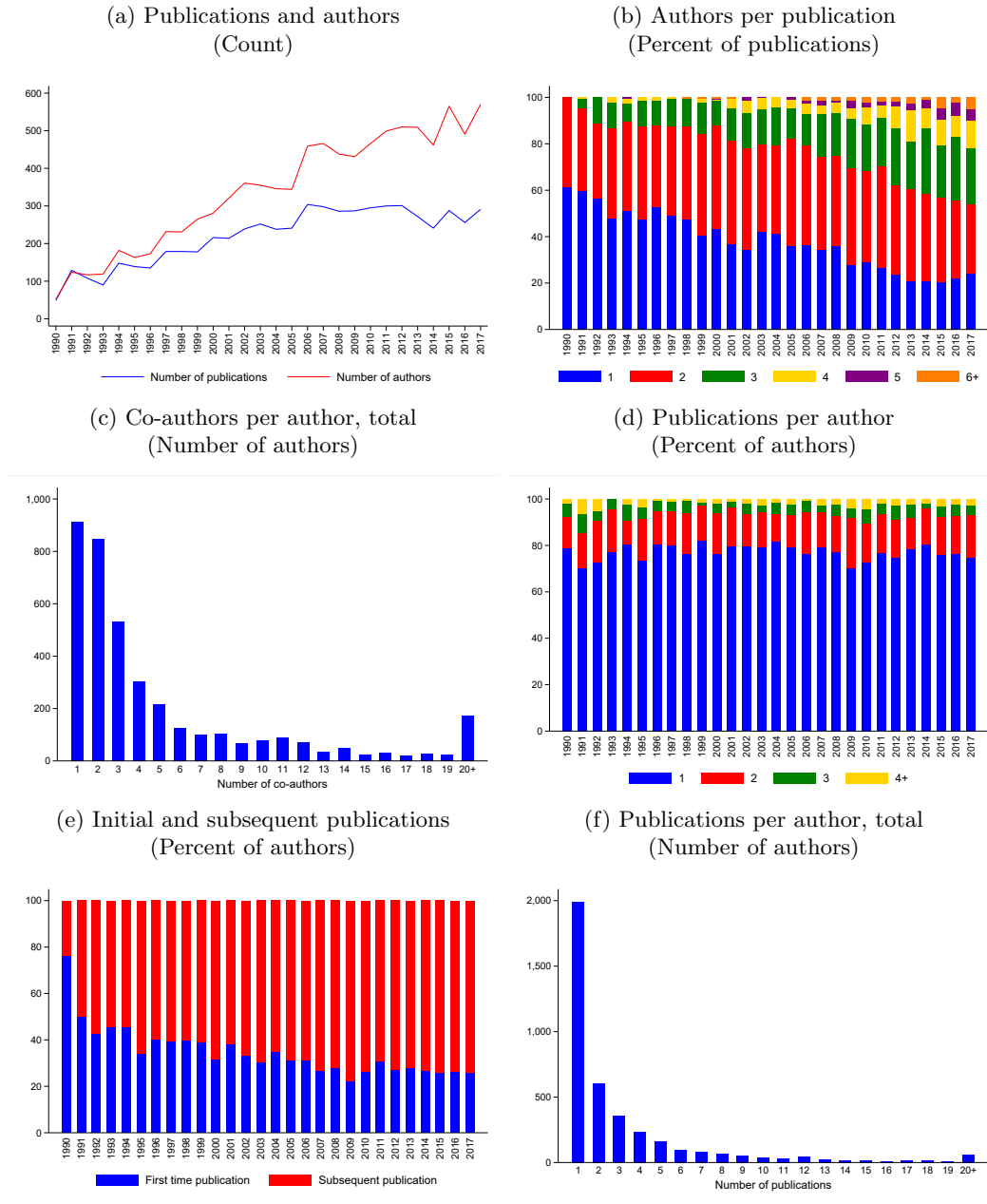
We now look at which author characteristics are associated with a higher propensity to do research, focusing on IMF staff. Figure 3 presents a set of charts in which, on the one hand, we consider the total share of authors by attribute in any given year (irrespective of whether the author published more than one paper that year); and, on the other hand, we account for the share of these authors in the IMF workforce. Due to the employee-level data availability, we cover only the period from 1999 to 2017.

As panel 3a shows, the majority of authors are economists (grades A11–A15), who account for over 80 percent of all publishing authors. The share of the support staff, which are primarily research assistants and other more senior specialists (grades A6–A10), rose to about ten percent since 1999, broadly compensated by a decline in the share of authors with managerial-level positions (grade B1–B5) from nearly 20 percent to ten percent. At the same time, for all three groups the share of employees who publish expressed as percent of the total workforce for the corresponding group increased over time (panel 3b). In 2017, about 20 percent of economists, seven percent of support staff, and 15 percent of managers (focusing on managers with an economic background) published at least one paper.

Around 40 percent of publishing authors nowadays work in area departments—African Department (AFR), Asia and Pacific Department (APD), European Department (EUR), Middle East and Central Asia Department (MCD), and Western Hemisphere Department (WHD)—with the other 60 percent mostly taken up by authors from functional departments—Research Department (RES), Strategy, Policy and Review Department (SPR), Statistics Department (STA), Fiscal Affairs Department (FAD), Money and Capital Markets Department (MCM), and Institute of Capacity Development (ICD). In recent years, the largest share of authors was from RES, followed closely by WHD (panel 3c). When looking at the number of publishing economists (i.e., combining non-managerial and managerial levels)—who account for the bulk of publishing authors—as percent of economists employed by the respective departments, RES stands out with the highest publication rate. Among area departments, there has been a decline since 2009 in the share of economists who publish in EUR, and an increase in WHD and APD (panel 3d).

The share of female authors gradually increased over time and reached about 30 percent in recent years (panel 3e), largely reflecting the increase in the share of women in the IMF workforce.

Figure 2: Collaboration Characteristics

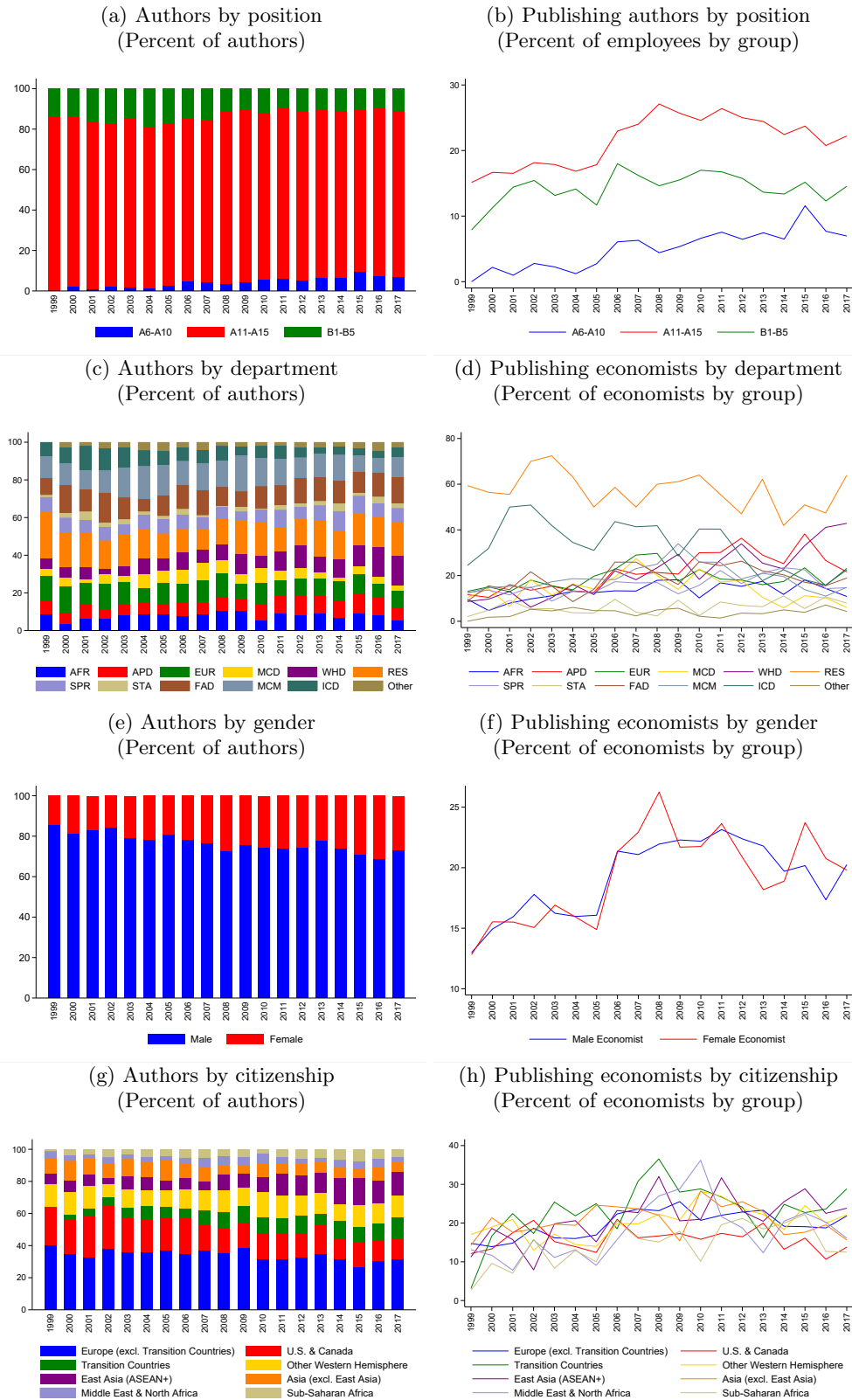


*Notes:* The panels in the figure present some key aspects of collaboration characteristics. In panel 2a, the number of authors represents unique authors, as some authors wrote multiple papers in the same year. Panels 2c and 2f show, respectively, the distribution of authors by the total number of co-authors (excluding researchers who wrote only single-authored papers) and the total number of publications released over the full sample period. Panel 2e shows, in any given year, the percentage of authors who published their first working paper vs. the percentage of authors who had written before (multiple papers released within the same year are treated as one). The downward trend is in part due to the fact that our dataset does not include information on publications written prior to 1990.

Indeed, once the shares of male and female economists in the workforce are taken into account, there are no systematic gender differences in the percent of economists who publish (panel 3f). In 2017, about 20 percent of both male and female economists published at least one paper. These findings compare favorably to those of Ductor et al. (2018) who find that, in per capita terms, women persistently published about 25 percent fewer articles than men in economics journals over 1996 to 2017.

We finally look at the region of citizenship of the authors. In 2017, slightly less than half of the publishing authors were from Europe and North America, yet their share in total authors has been decreasing over time. Two geographical regions largely compensated for this decline: transition countries and East Asia (panel 3g). Expressed as percent of the economists employed by the IMF from each corresponding region of citizenship, the publication rates for economists of different citizenship broadly tracked each other over time, with only slightly higher rates for the transition countries and East Asian regions in recent years (panel 3h).

Figure 3: Author Characteristics



*Notes:* The panels in the figure present some key aspects of author characteristics of IMF Working Papers. Panels on the left show the share of authors in specific groups relative to the total number of authors, irrespective of whether the authors published more than one paper. Panels on the right show the share of economists in specific groups who published relative to the total number of economists in those groups. In panel 3b, the set of A6–A10 employees is limited to positions with direct research duties based on the job title (research assistants, research officers, among others), while the set of A11–A15 employees (with non-managerial positions) and B1–B5 employees (with managerial positions) correspond to economists only. Panels 3d, 3f, and 3h focus on economists, combining non-managerial and managerial positions.

### 2.2.3 Author Pairs

Figure 4 illustrates some stylized facts for the pairs of authors, as defined in Section 2.1. The number of active pairs peaked at over 400,000 in 2011–2013 (panel 4a). As a reference, the IMF has around 3,000 employees. If all of these employees were active researchers, there would have been around 4.5 million active pairs ( $3,000 \times 2,999/2$ ). Since 2014, the number of active pairs has declined. This likely reflects a mechanical effect: researchers might have become inactive in recent years if they have not published any papers, yet they could again be re-classified as active if they publish a paper at a later date. On average, just over 0.1 percent of active pairs actually published together.

The great majority of author pairs were active for only one year (panel 4b), reflecting the relatively high frequency of one-time authors. Longer activity spans are much less common. Also, most author pairs that published, released only a single paper (panel 4c). In panels 4d and 4e, the total number of active pairs is broken down by author attributes: both authors are male, both are female, or authors are of different gender; both are economists, both are non-economists, or one author is an economist and the other is not; authors share the same region of citizenship, or not; and authors work in the same department, or not. For the rarer types of combinations, the number of active pairs is smaller (for example, there are fewer authors who work in the same department than the total pool of authors who work in different departments). As such, because of the smaller denominator, the publication rate by pairs in that group tends to be inflated.<sup>13</sup> With that important caveat in mind, women appear to collaborate relatively more with authors of the same sex than men do, economists collaborate more among themselves, and authors with the same region of citizenship or employed in the same department have higher propensities to collaborate.

Lastly, the vast majority of active pairs with at least one publication worked at the IMF for less than ten years. Junior-senior collaborations are also relatively common, whereas senior-senior collaborations are less frequent, likely because senior staff tend to have less time to devote to research (panel 4f).

### 2.2.4 Publication Length and Citations

Figure 5 summarizes the information on the number of pages and the number of citations per working paper. The number of pages ranged, on average, between 30 and 40 over the years (panel 5a), which is somewhat lower than for the average article in the top-tier economics journals (Card and DellaVigna, 2013). The median number of pages tends to be lower than the mean, indicating a right-skewed distribution due to some exceptionally long publications.

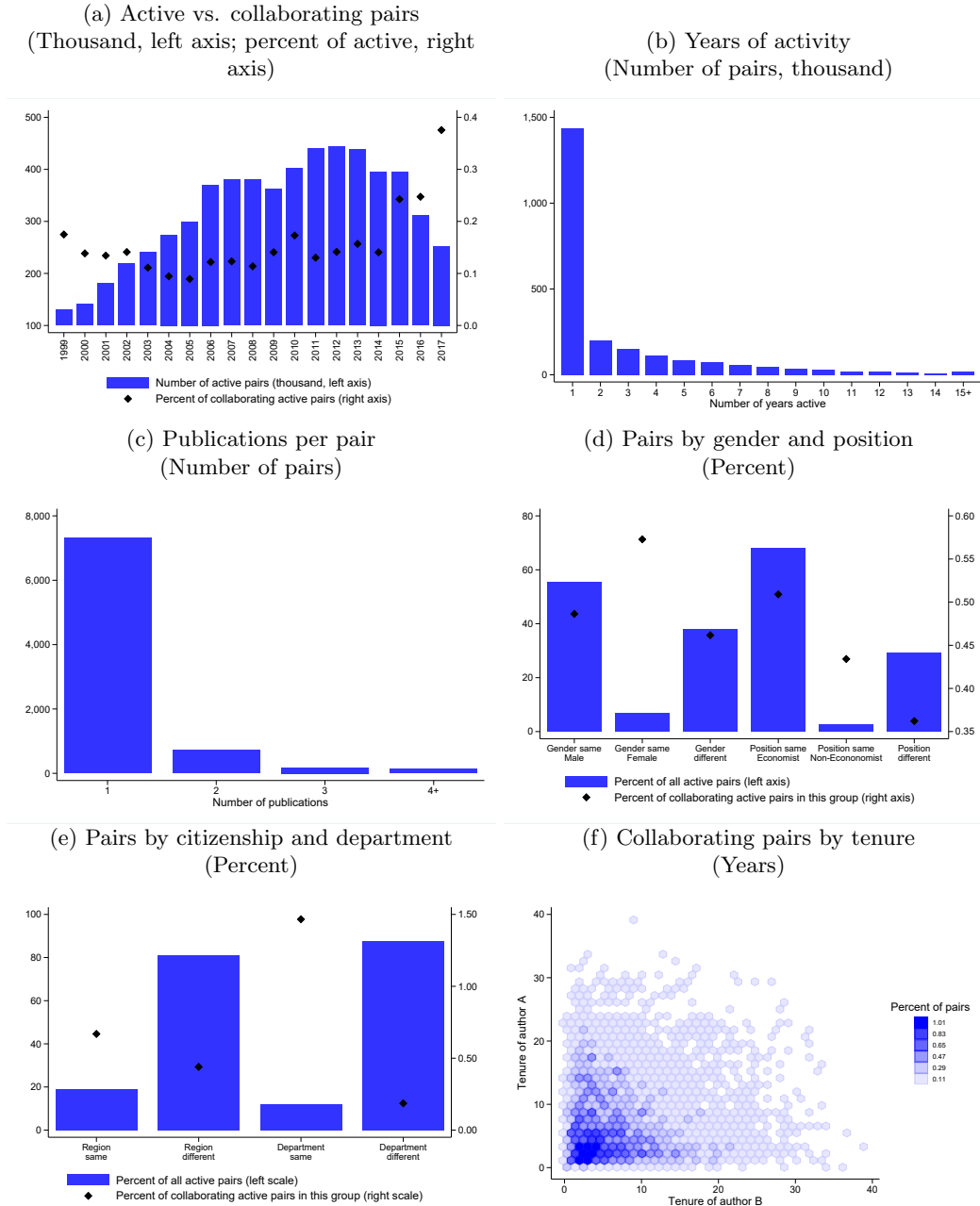
We then look at the cumulative number of citations as of January 2019 for all working papers.<sup>14</sup> To account for the fact that papers were published in different years, we annualize the number of citations using the number of years since publication. The average number of citations hovered between five and ten, with the median number of citations ranging between zero and five.

---

<sup>13</sup>For example, if just two authors share the same attribute and they collaborate, the percent of collaborating authors in that group would jump to 100 percent.

<sup>14</sup>For a more in-depth, comparative study of the citation record of IMF Working Papers, see Aizenman et al. (2011).

Figure 4: Pair Characteristics



Notes: The panels in the figure present some key aspects of author pairs' characteristics of IMF Working Papers. In panel 4a, the right axis shows active pairs that collaborated as a share of all active pairs. Panel 4c shows only the number of pairs that have at least one publication. In panel 4f, darker colors indicate a larger share of pairs with at least one publication.

The dispersion around the median, however, is large as there are some papers that appear to have been much more influential (panel 5b). The right-skewed distribution of citations has also been observed for economics papers more generally (Hamermesh, 2018; Hsieh et al., 2018).

The number of pages correlates with the number of authors (panel 5c). Publications with six or more authors are about 15 pages longer than single-authored publications, based on the median values. This pattern, however, is only partially reflected in the number of citations, as papers with three to four authors tend to get the highest number of citations (panel 5d).

There is no discernible difference in the number of pages for male and female authors. Even across departments, there are no striking differences; if anything, papers written by authors in RES and in MCM tend to be slightly longer (panel 5e). At the same time, while the number of citations for male and female authors is similar, papers written by authors working in RES tend to have noticeably higher citation counts (panel 5f).

### 2.2.5 Research Topics

We now turn to the topics covered by the IMF Working Papers. In Figure 6, we use the *Journal of Economic Literature* (JEL) classification codes to distinguish between different research areas.<sup>15</sup> Given the IMF's mandate, the majority of publications is unsurprisingly classified under codes E (Macroeconomics and Monetary Economics) and F (International Economics). These two codes account for about half of total publications. Adding codes G (Financial Economics) and H (Public Economics), about 75 percent of all publications is covered (panel 6a).

In general, working papers deal with topics related to more than one category (panel 6b). In recent years, about 80 percent of total publications reports more than one JEL letter, and 30 percent reports more than two JEL letters, with two letters being the most commonly used. Similarly, most authors use two JEL letters over the course of their active publishing years at the IMF (panel 6c). This, however, is in part due to the fact that the majority of the authors has only one publication. All in all, there is a significant number of authors who have used a wider range of JEL codes, indicating broad research interests.

We now look at the patterns emerging when more than one JEL code is used within a single publication (panel 6d). A heat map with the publication count for all bivariate combinations of JEL codes reveals that the most common codes E, F, G, and H are often used in combination. Other recurrent combinations are with code C (Mathematical and Quantitative Methods) and code O (Economic Development).

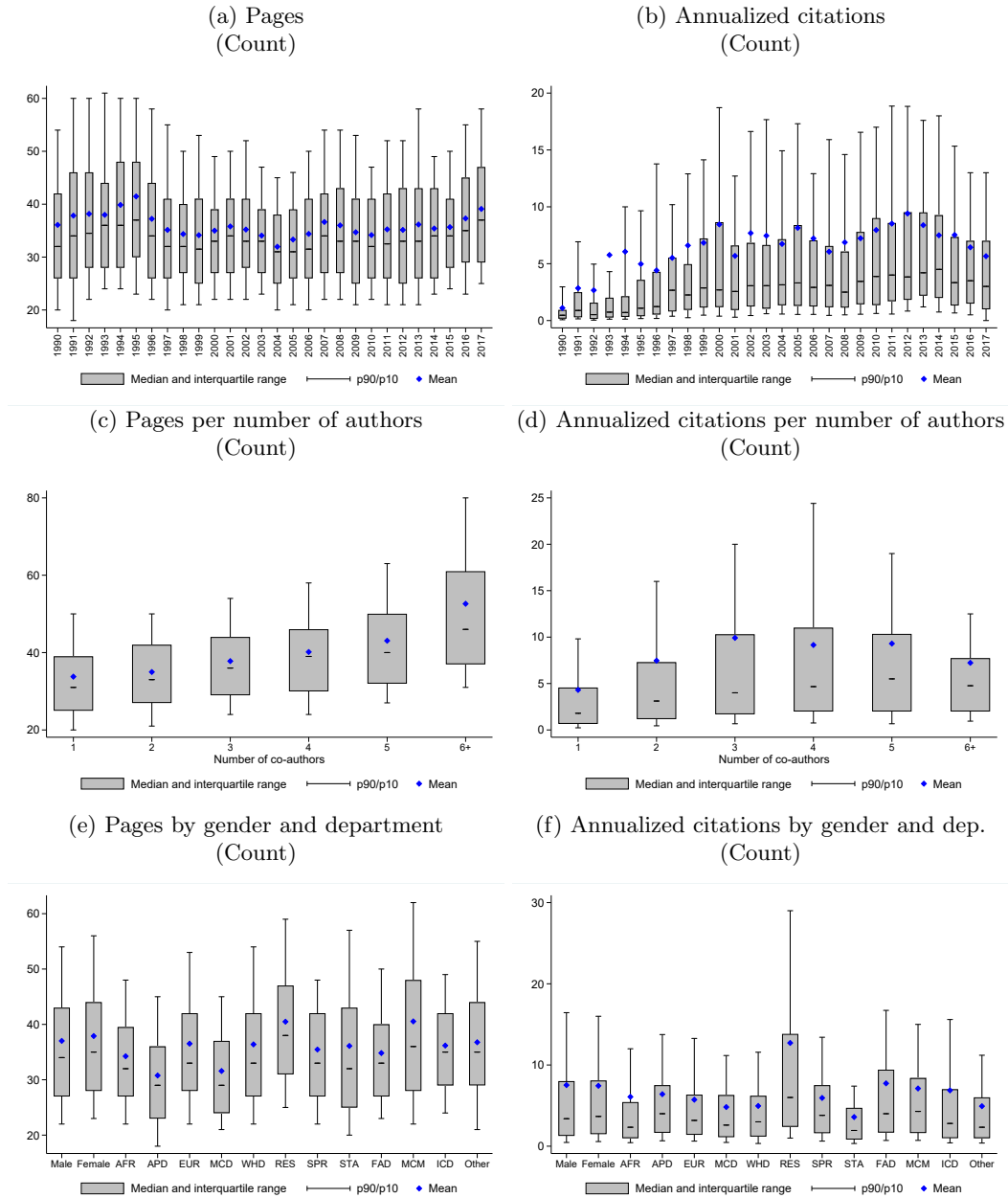
In recent years, increased attention has been dedicated to some topics that are less conventional for the IMF. For example, there has been an uptick in the number of papers written on cli-

---

<sup>15</sup>The JEL codes start with a letter and can be further disaggregated with up to three digits. For the purpose of this analysis, we use only the first letter of the JEL codes. These are A–General Economics and Teaching; B–History of Economic Thought, Methodology, and Heterodox Approaches; C–Mathematical and Quantitative Methods; D–Microeconomics; E–Macroeconomics and Monetary Economics; F–International Economics; G–Financial Economics; H–Public Economics; I–Health, Education, and Welfare; J–Labor and Demographic Economics; K–Law and Economics; L–Industrial Organization; M–Business Administration and Business Economics; N–Economic History; O–Economic Development, Innovation, Technological Change, and Growth; P–Economic Systems; Q–Agricultural and Natural Resource Economics; R–Urban, Rural, Regional, Real Estate, and Transportation Economics; Y–Miscellaneous Categories; Z–Other Special Topics.



Figure 5: Publication Lengths and Citations



*Notes:* The panels in the figure present some key aspects of publication length and citations of IMF Working Papers. The statistics in panels 5e and 5f are calculated as simple averages over authors and working papers. For example, if a paper has two authors with opposite genders and  $x$  citations, and another paper is written by one female author and has  $y$  citations, then the average number of citations for female authors would be  $(x + y)/2$  and the average number of citations for male authors would be  $x$ .

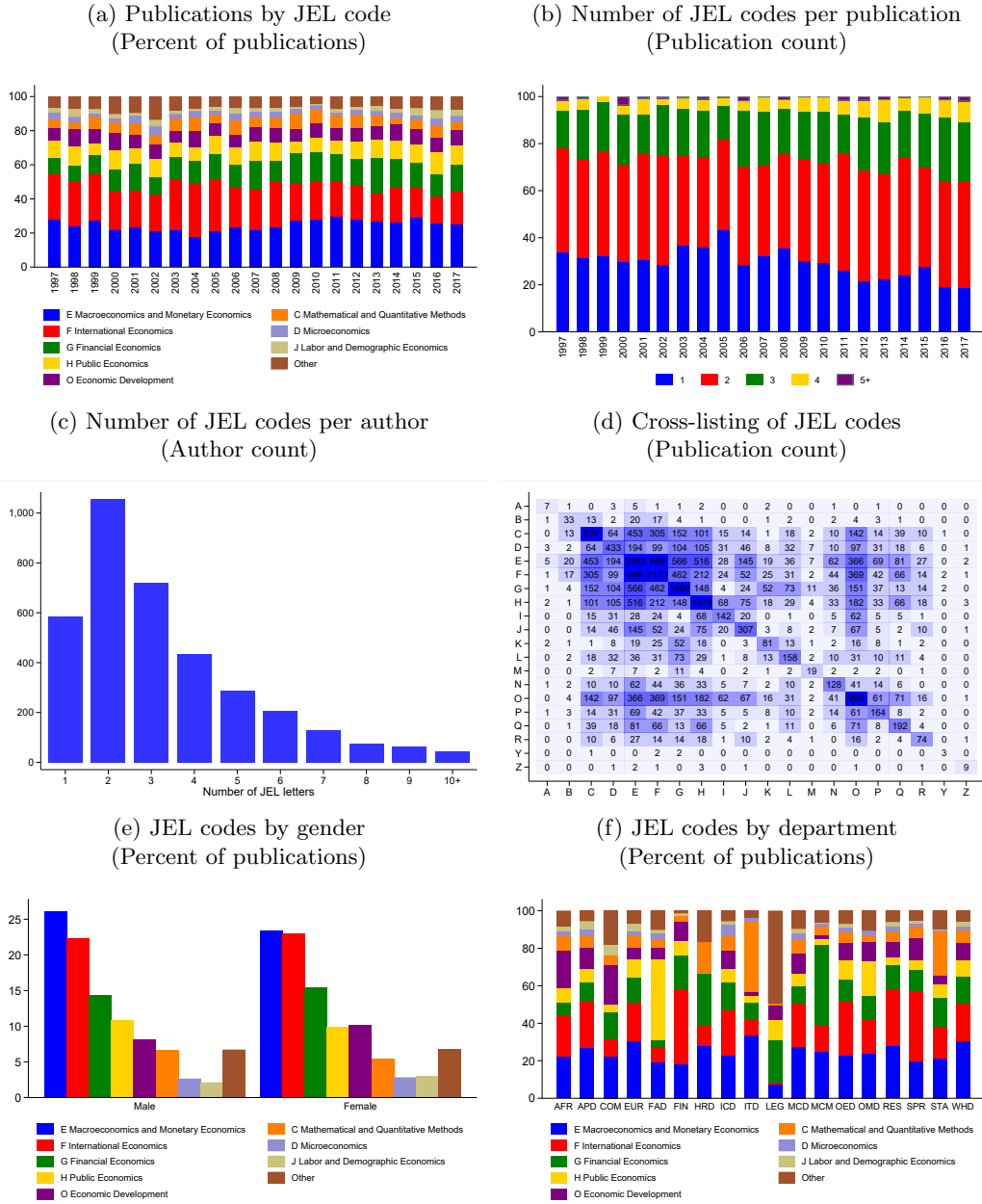
mate change (JEL code Q54) since 2012, and on gender (J16) since 2015, albeit from a very low base.<sup>16</sup>

Male and female authors publish within the same JEL codes in broadly equal proportions (panel 6e). If anything, female authors publish marginally fewer papers under code E (Macroeconomics and Monetary Economics), and somewhat more across most other JEL codes. Some differences, however, are evident across departments (panel 6f). The heterogeneity is roughly in line with each department’s mandate and specialization. For example, researchers in FAD publish more papers within code H (Public Economics), while researchers in MCM publish more papers within code G (Financial Economics).

---

<sup>16</sup>In several IMF documents, including recent Annual Reports, climate change and gender are identified as (macro-critical) “emerging issues”, often together with inequality and fintech. Inequality (JEL codes D63, I14, and I24) has been a recurrent IMF Working Paper topic since the mid-1990s. Conversely, there appear to be only very few working papers that deal with fintech (which falls under the broader JEL code G2).

Figure 6: Publication Topics



Notes: The panels in the figure present some key aspects of publication topics of IMF Working Papers. In the legend of panel 6a, JEL codes are ordered by the share size. In panels 6a, 6e, 6f, publications with multiple JEL (lette) codes are assigned proportionally to each of the corresponding code (for example, a 50/50 split for a paper that uses two codes). These shares are then summed up to measure the total amount of research dedicated to each JEL code. The diagonal in panel 6d shows the number of publications that use the corresponding JEL code. The same publication could be cross-listed with other JEL codes, appearing off-diagonal. For a full list of department names in panel 6f see “International Monetary Fund Organization Chart” (<https://www.imf.org/external/np/obp/orgcht.htm>, accessed June 2020).

### 3 Co-Authorship Network

Having described the characteristics of the publications, of the authors, and of the author pairs, we now consider the broader network in which these are embedded. Indeed, the literature on co-authorship in economics and other fields has demonstrated that such networks matter a great deal, notably in facilitating new research collaborations and in raising the quality of research output (see Section 1). In the following sections we study the co-authorship network both at the level of the individual authors and at the level of the different IMF departments and external institutions with which these authors are affiliated. Making use of the relatively long time dimension of our dataset, we also examine how the network structure has evolved over time.

#### 3.1 Author-Level Network

Figure 7 shows the author-level network of all IMF Working Papers published between 1990 and 2017, whereby the nodes represent individual authors and the undirected edges joining them represent co-authorship of a publication. The size of the nodes is proportional to the number of publications an author wrote over the entire period, including single-authored papers (which do not give rise to an edge). The edge width is proportional to the number of repeated collaborations between two authors (including multiple publications within the same year). Note that a publication with three authors implies three collaborations and thus three edges; a publication with four authors results in six edges, and so on.<sup>17</sup> Node colors refer to authors’ affiliations, which are divided into the six categories (beyond the IMF) mentioned in Section 2.1 (i.e., universities, other international organizations, central banks, government agencies, non-profit organizations, and private organizations). Since people may switch between affiliations over time, the node color denotes the mode of an author’s affiliation category over the years in which she/he published.

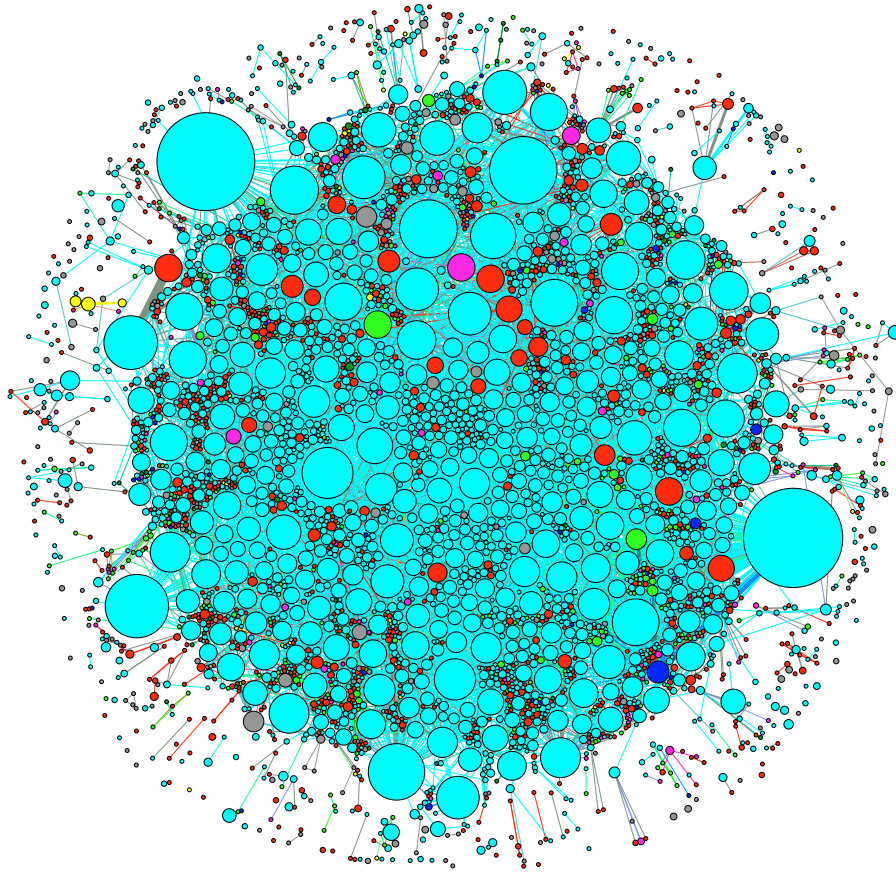
In Table 1 we provide some summary statistics often used to describe networks. As shown in column (1), the full IMF Working Papers network counts 3,918 nodes and 9,027 edges, resulting in an average “degree”, or number of co-authors per researcher, of about 4.6. What is also clear from Figure 7 is that most authors are in some way connected to each other and that the network has relatively few isolated authors or author groups. In fact, the network has a so-called “giant component”, i.e. the largest group of nodes that are all directly or indirectly connected to each other, that covers 3206 nodes, or almost 82 percent of all authors in the entire network. In that sense, research at the IMF does not seem to consist of distinct “islands” of researchers. The average “distance”, or length of the shortest path between two authors in the giant component of the network, is about five. The average “clustering coefficient” is a measure of the overlap in co-authorship and defined as the fraction of an author’s collaborators that are themselves also co-authors (averaged over all authors with a minimum degree of two). Here it equals more than 0.6.

To put these numbers in perspective, column (5) of Table 1 lists the corresponding network statistics for the network of authors of all articles published over 1990–1999 in journals indexed in EconLit (Goyal et al., 2006). Compared to the EconLit network, our IMF Working Papers

---

<sup>17</sup>The number of collaborations involved in a paper is given by  $n(n - 1)/2$ , where  $n$  is the number of authors.

Figure 7: Full Author-Level Network, 1990–2017



*Notes:* The figure illustrates the full author-level network based on IMF Working Papers co-authorship over 1990–2017. The nodes represent authors and the edges represent publication co-authorship. The node size is proportional to the number of publications by an author (including single-authored papers). The edge width is proportional to the number of publication collaborations between two authors. The node colors are according to the modes of authors' affiliation category: light blue denotes the IMF, pink denotes other international organizations, red denotes universities, light green denotes central banks, dark green denotes government agencies, yellow denotes non-profit organizations, dark blue denotes private organizations, and grey denotes N/A (for authors that were only active prior to 1997 or that have no unique affiliation category mode).

Table 1: Network Statistics

|                                     | IMF Working Papers |                  |                  |                  | EconLit articles |
|-------------------------------------|--------------------|------------------|------------------|------------------|------------------|
|                                     | 1990–2017<br>(1)   | 1990–1999<br>(2) | 2000–2009<br>(3) | 2010–2017<br>(4) | 1990–1999<br>(5) |
| Total nodes (authors)               | 3,918              | 923              | 1,891            | 2,192            | 81,217           |
| Total edges (co-authorship)         | 9,027              | 1,257            | 4,081            | 6,058            | 67,897           |
| Average degree                      | 4.608              | 2.724            | 4.316            | 5.527            | 1.672            |
| Size of giant component             | 3,206              | 594              | 1,598            | 1,916            | 33,027           |
| Share of giant component (percent)  | 81.8               | 64.4             | 84.5             | 87.4             | 40.7             |
| Average distance in giant component | 4.961              | 5.132            | 4.878            | 4.640            | 9.470            |
| Average clustering coefficient      | 0.614              | 0.443            | 0.504            | 0.613            | 0.157            |

*Notes:* The table reports network statistics for the co-authorship network of IMF Working Papers published over 1990–2017 and sub-periods thereof, as well as for the co-authorship network of all journal articles indexed in EconLit according to Goyal et al. (2006).

network is much smaller, especially taking into account the nearly three decades over which it stretches. Conversely, the IMF Working Papers network is considerably more integrated, as evident from the much greater average degree, coverage of the giant component, and average clustering coefficient. There are arguably different explanations for the relatively high integration of the IMF Working Papers network, besides the long time frame (which increases chances that researchers collaborate at some point). Importantly, we focus on the publications of one institution, where potential collaborators may get to know each other more easily (even just because staff mostly works in the same premises) and researchers cover a smaller set of topics than in the field of economics more generally.<sup>18</sup> Moreover, by definition clustering increases with the number of multi-authored publications, which account for a non-negligible share of IMF Working Papers, especially in more recent years.<sup>19</sup>

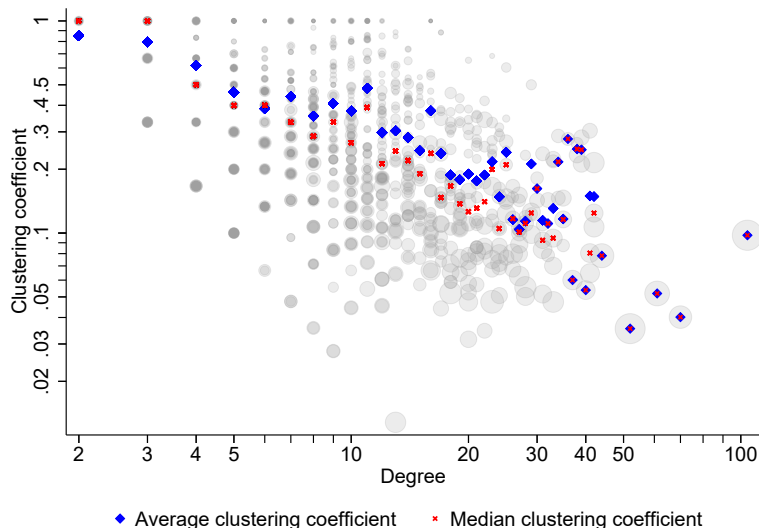
Taken together and in line with the findings for collaborations in economics more generally (Goyal et al., 2006) and in other fields of research (Newman, 2001), the network measures in Table 1 indicate that the IMF Working Papers network exhibits clear “small world” properties. That is, the network consists of a large set of authors with few direct co-authors (relative to all possible collaborations); the network is nonetheless highly integrated, as the majority of authors are in some way connected to each other, mostly through short chains of co-authors; and the extent of overlap in co-authorship is very high (relative to a network where collaborations would be purely random). As Goyal et al. (2006) and Hsieh et al. (2018) show, such patterns typically derive from a network architecture of interlinked “star” authors, a set of researchers that collaborate with many others, most of which are not direct co-authors themselves. These stars effectively act as connectors of different clustered parts of the network. Additional evidence for such a set-up comes

<sup>18</sup>There is a large literature showing that, by reducing search costs, physical proximity (even if temporary, during conferences, workshops, information sessions, etc.) significantly increases the likelihood of initiating scientific collaboration and co-authorship. See, e.g., Boudreau et al. (2017), Catalini (2018), Campos et al. (2018), and Chai and Freeman (2019).

<sup>19</sup>Note that, for technical reasons, Goyal et al. (2006) exclude from their analysis of the EconLit network all articles with four or more authors, which represents a reported 1.6 percent of the full sample.

from the observed negative relation between individual authors’ degree and their local clustering coefficient in Figure 8. Whereas the many researchers with only two or three co-authors have on average (at the median) around 80 percent (100 percent) of their co-authors collaborating with each other, the fewer researchers with over 20 co-authors seldom have clustering coefficients in excess of 0.2.

Figure 8: Author Degrees vs. Clustering Coefficients



*Notes:* The figure plots individual authors’ degree against their clustering coefficient for all authors that published IMF Working Papers over 1990–2017 and had a minimum degree of two. It also shows the average and median clustering coefficients per degree. Both axes have a logarithmic scale. Authors with clustering coefficients of zero are not shown but are taken into account in the calculation of the averages and medians. The size of the dots is proportional to the total number of publications by the author. Darker shades of grey correspond to more authors having the same degree-clustering coefficient combination.

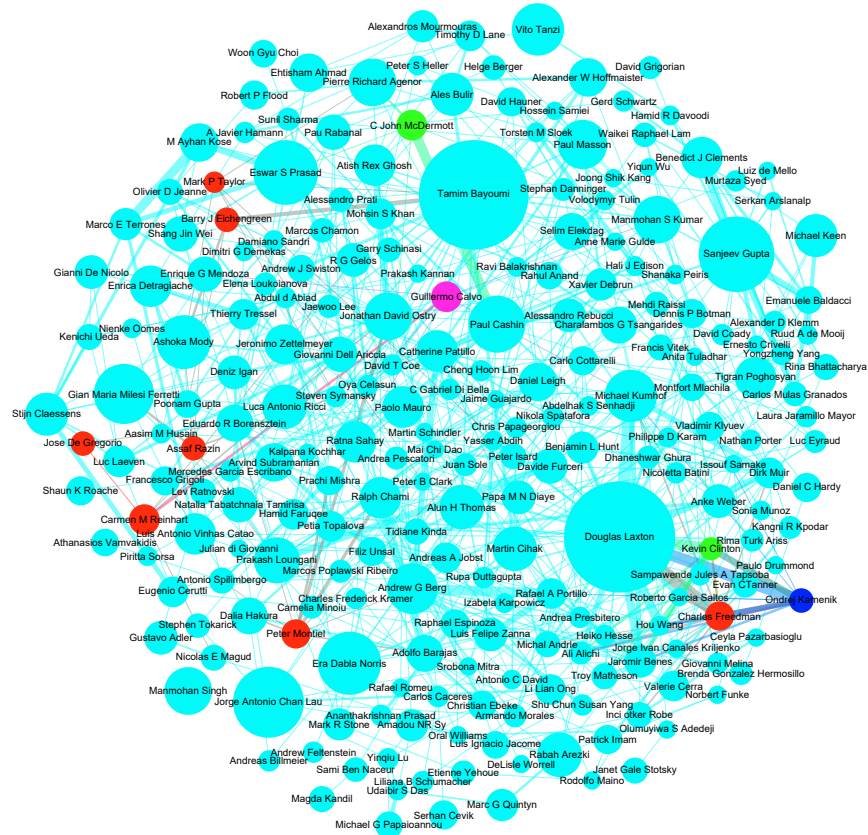
When we consider authors’ affiliation in Figure 7, it is immediately evident that the network is dominated by IMF staff, unsurprisingly so. By far, IMF staff account for the largest number of unique authors (2,125 nodes) and for the largest number of publications by those authors (node size). Academics, which include resident and visiting scholars at the IMF, come a distant second (964 nodes), followed by central bankers (258), and staff from other international organizations (109). Authors affiliated with government agencies (72), private companies (52), or non-profits (35) occupy a much less prominent role in the network.

To make the structure of the network more visually tractable and to better grasp how it looks at its core, in Figure 9 we restrict the network to those authors that contributed to at least ten publications and that collaborated with at least ten different co-authors (in other words, they have a minimum degree of ten) over 1990–2017. In this “core” network of 233 authors it is even clearer that the most prolific and best-connected authors are all IMF staff.<sup>20</sup> The few non-IMF

<sup>20</sup>Note that most nodes in Figure 9 are connected by less than ten edges, as collaborations with “non-core” authors are counted but not shown.

authors present in the core network are predominantly academics that have been regularly visiting or were at some point employed at the IMF.

Figure 9: Core Author-Level Network, 1990–2017



*Notes:* The figure illustrates the core author-level network based on IMF Working Papers co-authorship over 1990–2017. The sample includes authors who contributed to at least ten publications and collaborated with at least ten different co-authors (in other words, authors with a minimum degree of ten). The nodes represent authors and edges represent publication co-authorship. The node size is proportional to the number of publications by an author (including single-authored papers). The edge width is proportional to the number of publication collaborations between two authors. The node colors are according to the nodes of authors' affiliation category: light blue denotes the IMF, pink denotes other international organizations, red denotes universities, light green denotes central banks, dark blue denotes private organizations.



## 3.2 Institution-Level Network

Another level at which the network can be studied is that of the institutions to which (non-IMF) authors are affiliated. Figure 10 collapses all individual co-authorship relations over 1997–2017 by institution (whereby node colors are again according to affiliation category). It is evident that IMF staff has been collaborating with researchers from many different institutions (639 in total, excluding self-employed consultants), mostly universities (380 institutions) and central banks (77). The node size in Figure 10 is proportional to the number of publications to which at least one author affiliated with the institution in question contributed, whereas the edge width is now proportional to the number of publication collaborations between the two respective institutions that the edge connects.<sup>21</sup> There is clear heterogeneity in the intensity of collaborations (i.e., the intensive margin) at the institutional level. Collaborations go from a one-time joint publication to repeated co-authorship with the same authors from a particular institution or co-authorship with many different authors affiliated with the same institution. Very few papers are written solely by authors with a non-IMF affiliation (i.e., nodes not connected to the central IMF node).

In Figure 11 we again consider the “core” of the network, this time by focusing on the institutions that contributed to at least ten different publications. The single most important collaboration at the institutional level is the World Bank (accounting for 98 distinct publications and 83 distinct authors). Academic collaborators hail mostly from world-renowned universities in the United States and the United Kingdom. A limited number of other universities located in Europe, Australia, Canada, and Israel are also found in this core network. We observe no academic institutions from emerging market and developing economies.<sup>22</sup> Central banks include the Federal Reserve Board, the European Central Bank, and a handful of (again mostly advanced economy) national central banks. OG Research, a firm specialised in economic and financial modelling, is the only private company found in the core (with more than ten publication involvements but just three authors). Besides the World Bank, also the Organisation for Economic Co-operation and Development and the Inter-American Development Bank contributed to about 20 publications each. Government agencies and non-profits are not represented.<sup>23</sup>

To help in understanding the structure of the co-authorship network *within* the IMF, Figure 12 disaggregates, where the data allows it, the affiliation of IMF authors by department (while collapsing non-IMF affiliations into the same six categories as before) for the period 1999–2017. Unsurprisingly, and in line with the findings of Section 2.2, RES is the most prominent IMF department in terms of publication (co-)authorship, even though, overall, academics contributed to more publications than RES staff (1,308 vs. 1,077 publications). RES is followed, at a distance, by MCM, FAD, and EUR (with involvements in more than 600 publications each). Across the different departments, RES has the most collaborations with universities, central banks, and interna-

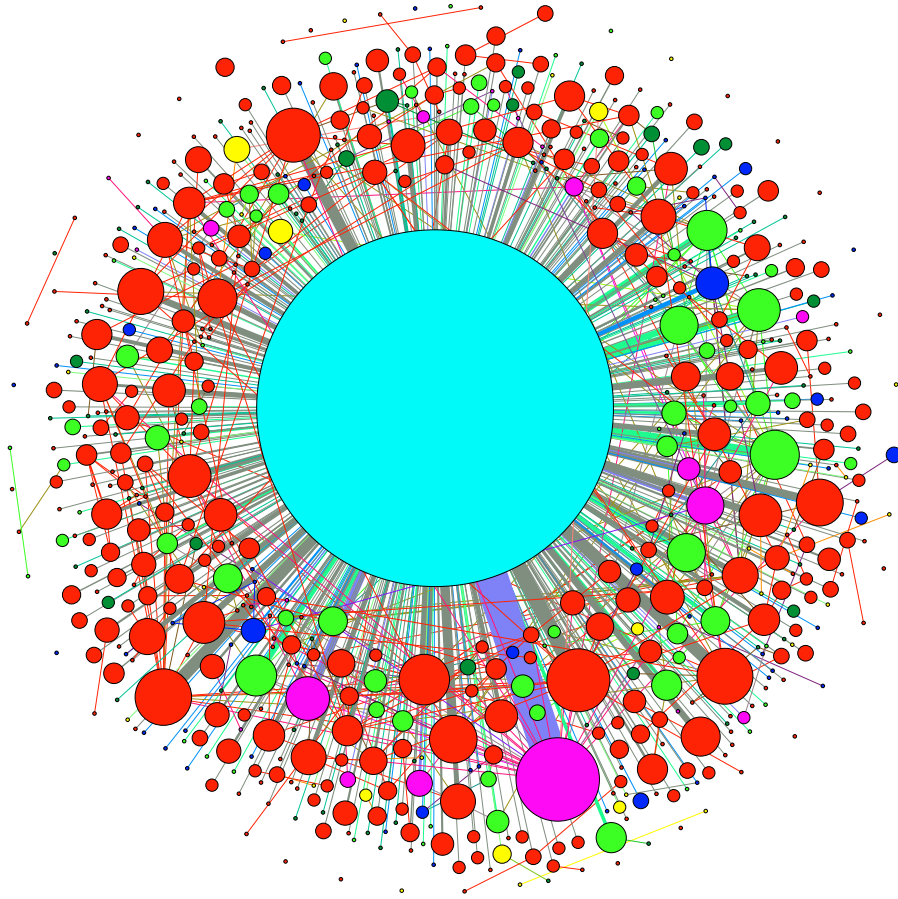
---

<sup>21</sup>Contributions by multiple authors of the same institution to a single publication and collaborations by the same pair of institutions within a single publication are counted only once in the calculation of the node size and edge width.

<sup>22</sup>With its involvement in seven publications, the Argentinean Universidad Torcuato Di Tella is the highest-ranked university located in an emerging market economy.

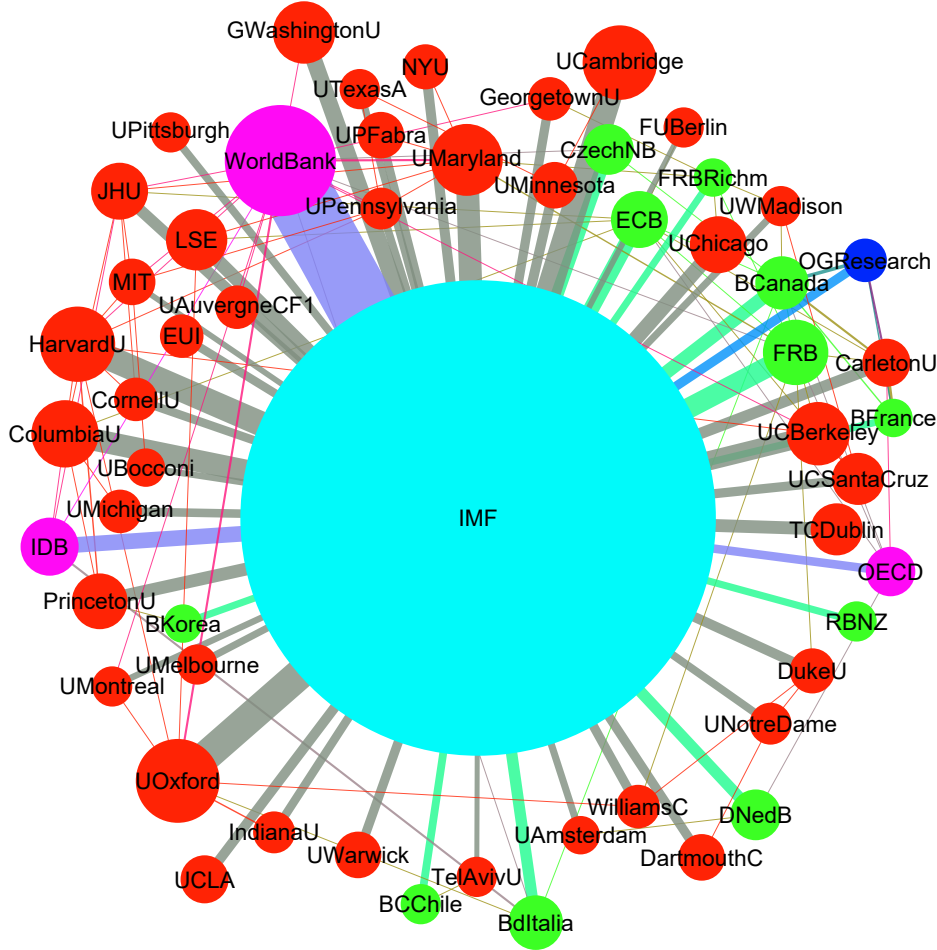
<sup>23</sup>The U.S. Treasury (involved in six publications) and India’s National Institute of Public Finance and Policy (eight publications) are the highest-ranked government agency and non-profit organization, respectively.

Figure 10: Full Institution-Level Network, 1997–2017



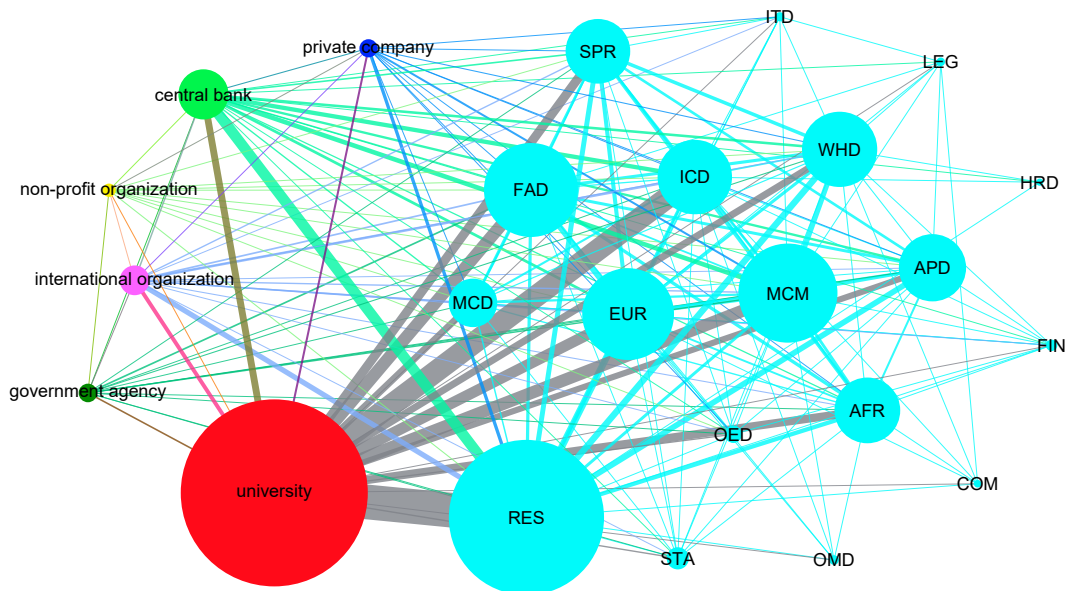
*Notes:* The figure illustrates the full institution-level network based on IMF Working Papers co-authorship over 1997–2017. The nodes represent institutions and the edges represent publication co-authorship by authors affiliated with the respective institutions (publication collaborations between authors of the same institution are discarded). The node size is proportional to the number of publications to which at least one author affiliated with the institution contributed (contributions by multiple authors of the same institution to a single publication are counted only once). The edge width is proportional to the number of publication collaborations between the two respective institutions (collaborations by the same pair of institutions within a single publication are counted only once). The node colors are according to institutional category: light blue denotes the IMF, pink denotes other international organizations, red denotes universities, light green denotes central banks, dark green denotes government agencies, yellow denotes non-profit organizations, dark blue denotes private organizations.

Figure 11: Core Institution-Level Network, 1997–2017



*Notes:* The figure illustrates the core institution-level network based on IMF Working Papers co-authorship over 1997–2017. The sample includes institutions that contributed to at least ten publications. The nodes represent institutions and the edges represent co-authorship by authors affiliated with the respective institutions (publication collaborations between authors of the same institution are discarded). The node size is proportional to the number of publications to which at least one author affiliated with the institution contributed (contributions by multiple authors of the same institution to a single publication are counted only once). The edge width is proportional to the number of publication collaborations between the two respective institutions (collaborations by the same pair of institutions within a single publication are counted only once). The node colors are according to institutional category: light blue denotes the IMF, pink denotes other international organizations, red denotes universities, light green denotes central banks, dark blue denotes private organizations.

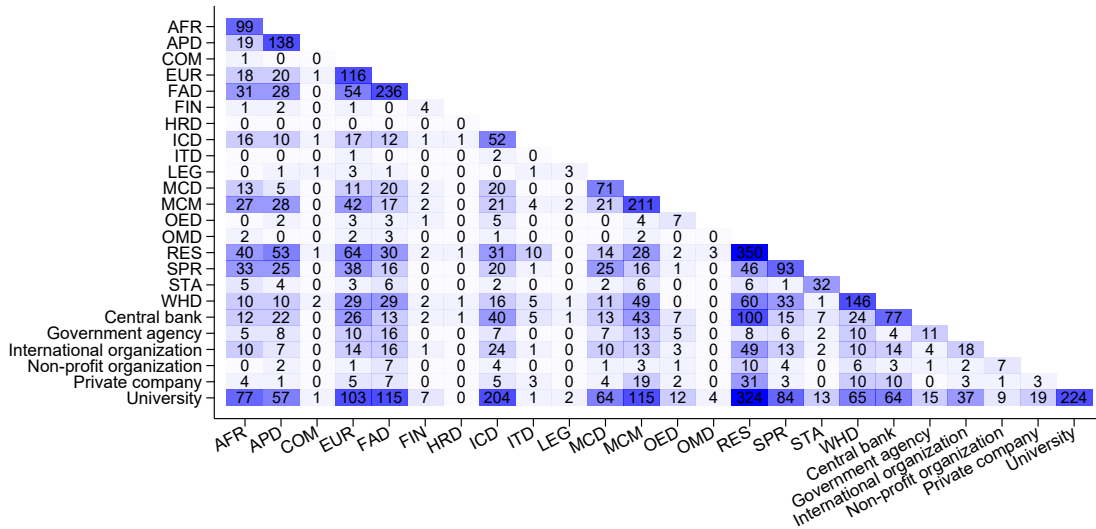
Figure 12: Full Network of IMF Departments and Other Institutions, 1999–2017



*Notes:* The figure illustrates the full network of IMF departments and other institutions based on IMF Working Papers co-authorship over 1990–2017. The nodes represent IMF departments (in light blue) and other institutions grouped by category (other colors), and the edges represent publication co-authorship by authors affiliated with the respective departments or institutional categories (collaborations between authors of the same department or institutional category are discarded). The node size is proportional to the number of publications to which at least one author affiliated with the department or institutional category contributed (contributions by multiple authors of the same department or institutional category to a single publication are counted only once). The edge width is proportional to the number of publication collaborations between the two respective departments or institutional categories (collaborations by the same pair of departments or institutional categories within a single publication are counted only once). 186 authors affiliated with the IMF for which no information on the department was available are excluded.

tional organizations. Within the IMF, RES collaborates mostly with EUR, WHD, APD, and SPR, but less so than with universities or central banks. Several IMF departments other than RES, notably ICD, MCM, FAD, EUR, and SPR, also collaborate more with universities than with other departments. Collaboration across area departments is relatively limited. An alternative way of visualising the foregoing is by means of a heat map. Figure 13 quantifies the number of publications by pair of departments or institutional categories and further demonstrates that research collaborations within the same department are more common than co-authorship across different departments.

Figure 13: Heat Map of Collaborations between IMF Departments and Other Institutions, 1999–2017



Notes: The heat map displays the numbers of publication collaborations between each pair of departments or institutional categories (collaborations by the same pair of departments or institutional categories within a single publication are counted only once). The numbers on the diagonal refer to collaborations between authors of the same department or institutional category. Darker colors correspond to larger absolute numbers of publication collaborations. 186 authors affiliated with the IMF for which no information on the department was available are excluded.

### 3.3 Evolution of the Network

We now further explore the time dimension of our dataset by looking at how the co-authorship network and some of its key attributes have evolved over time. Figure 14 graphically compares the author-level network over three decade-long periods: 1990–1999, 2000–2009, and 2010–2017. The three panels of Figure 14 represent the cumulative networks of authors that contributed to IMF Working Papers published over each of those respective periods. For 1990–1999, we discard author affiliations, which are only available from 1997 onward (see Section 2.1).

The network has grown larger and denser over time. As suggested by the results in Table 1, the number of authors more than doubled from just over 900 during 1990–1999 to almost 2,200 during 2010–2017. Meanwhile, the number of co-authorship links nearly quintupled between the first and the third period, resulting in an increase of the average degree from 2.7 to 5.5. During the 1990s the giant component already covered 64 percent of the authors publishing during that period, but this further increased to almost 85 percent in the 2000s and even more than 87 percent in the 2010s. The giant component also became more integrated, with an average distance progressively declining from 5.1 to 4.6 between the 1990s and 2010s. In addition, the overlap in co-authorship, as measured by the average clustering coefficient, rose considerably. Again, part of the increase in the the degree of clustering can be ascribed to the higher share of multi-authored publications in recent years.

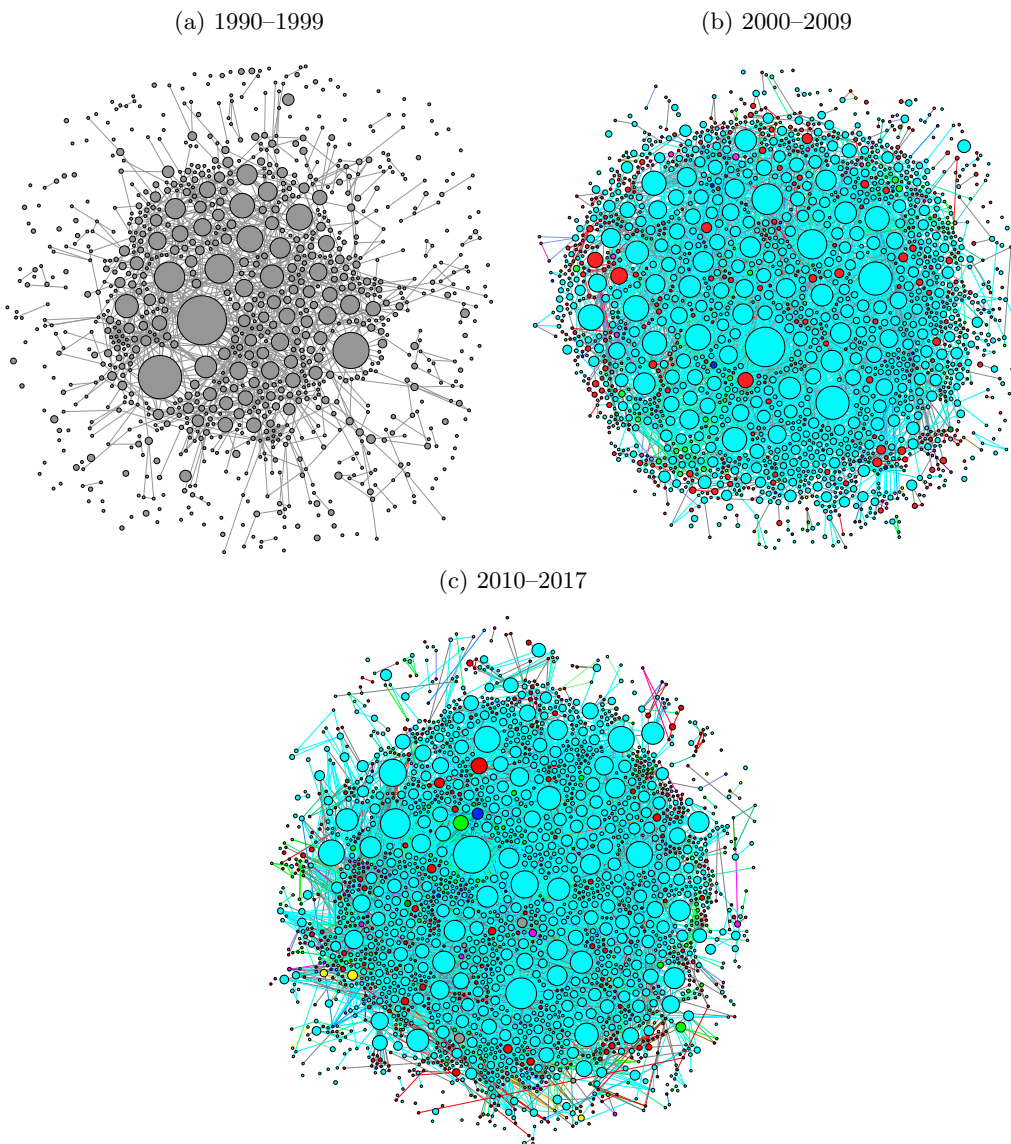
To shed further light on the evolution of some of the key attributes of our network over time, Figure 15 plots their distributions for the three sub-periods. Panel 15a shows the right-skewed nature of the distribution of authors’ degrees. In the 1990s, most authors had between zero and two co-authors, one being the mode of the distribution, and only very few authors with more than five co-authors. In the 2000s and then in the 2010s, the right tail of the distribution became fatter—even though the mode of the degree remained between one and two—suggesting that the noticeable increase in the average degree in Table 1 is mostly due to already well-connected authors further increasing the number of their collaborations.

Panel 15b demonstrates that the distribution of the clustering coefficient is even more heavily skewed to the right. The mode of the distribution remained close to zero over the sample period, but the right tail again became fatter over time. Interestingly, the distribution for the clustering coefficient presents a bump at the end of the right tail, indicating that some authors have most of their co-authors connected to each other. This is in part the result of authors that have just a single publication with two or more co-authors.

Panel 15c shows the distribution of the distance between authors, measured only for unconnected authors (i.e., excluding direct links between authors). The minimum value is two for those authors that share a common co-author. From the 1990s to the 2010s, the mode of the distribution moved from five to four and the distribution became tighter around the mode. This confirms that unconnected authors in the network became closer to each other as time went by.

Panel 15d looks at another indicator of network integration, the number of shortest paths that connect two authors in the network. In this case, unconnected authors have a value of zero. The mode of the density moved from zero to one by the 2010s. Yet again, the right tail of the distri-

Figure 14: Evolution of 10-year Cumulative Author-Level Network

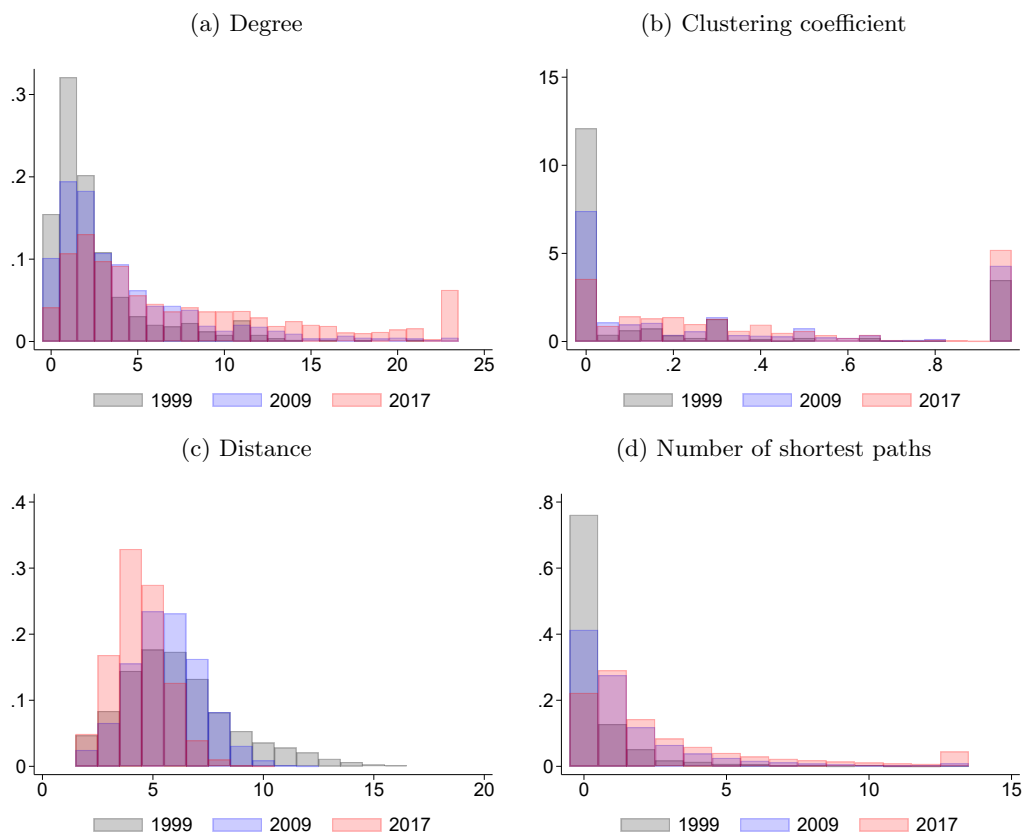


*Notes:* The figure illustrates the evolution of the 10-year cumulative author-level network based on IMF Working Papers co-authorship over 1990–2017. The nodes represent authors and edges represent co-authorship. The node size is proportional to the number of working papers written by an author (including single-authored papers) during the respective period: 1990–1999 (panel 14a), 2000–2009 (panel 14b), and 2010–2017 (panel 14c). Edge width is proportional to the number of collaborations between two authors during the respective period. In panels 14b and 14c, the node colors are according to the modes of authors' affiliation category over the respective period: light blue denotes the IMF, pink denotes other international organizations, red denotes universities, light green denotes central banks, dark green denotes government agencies, yellow denotes non-profit organizations, dark blue denotes private organizations, grey denotes N/A (for authors that have no unique affiliation category mode). Panel 14a discards affiliation categories, as information on affiliations is only available from 1997 onward.

bution became fatter, indicating that authors now have more equally-distant paths to reach other authors than previously.

If we look at author affiliation categories over time in Figure 14, we observe that IMF staff dominated the network during the second and third period (and most probably also the first period). In both periods, IMF authors represented just over 60 percent of all nodes. From the 2000s to the 2010s there was a slight increase in the diversity of external collaborations, away from academics (which remained, by far, the most important external category) towards more central bankers and researchers affiliated with non-profits (mostly think tanks).

Figure 15: Distributions of Network Statistics  
(Density units)



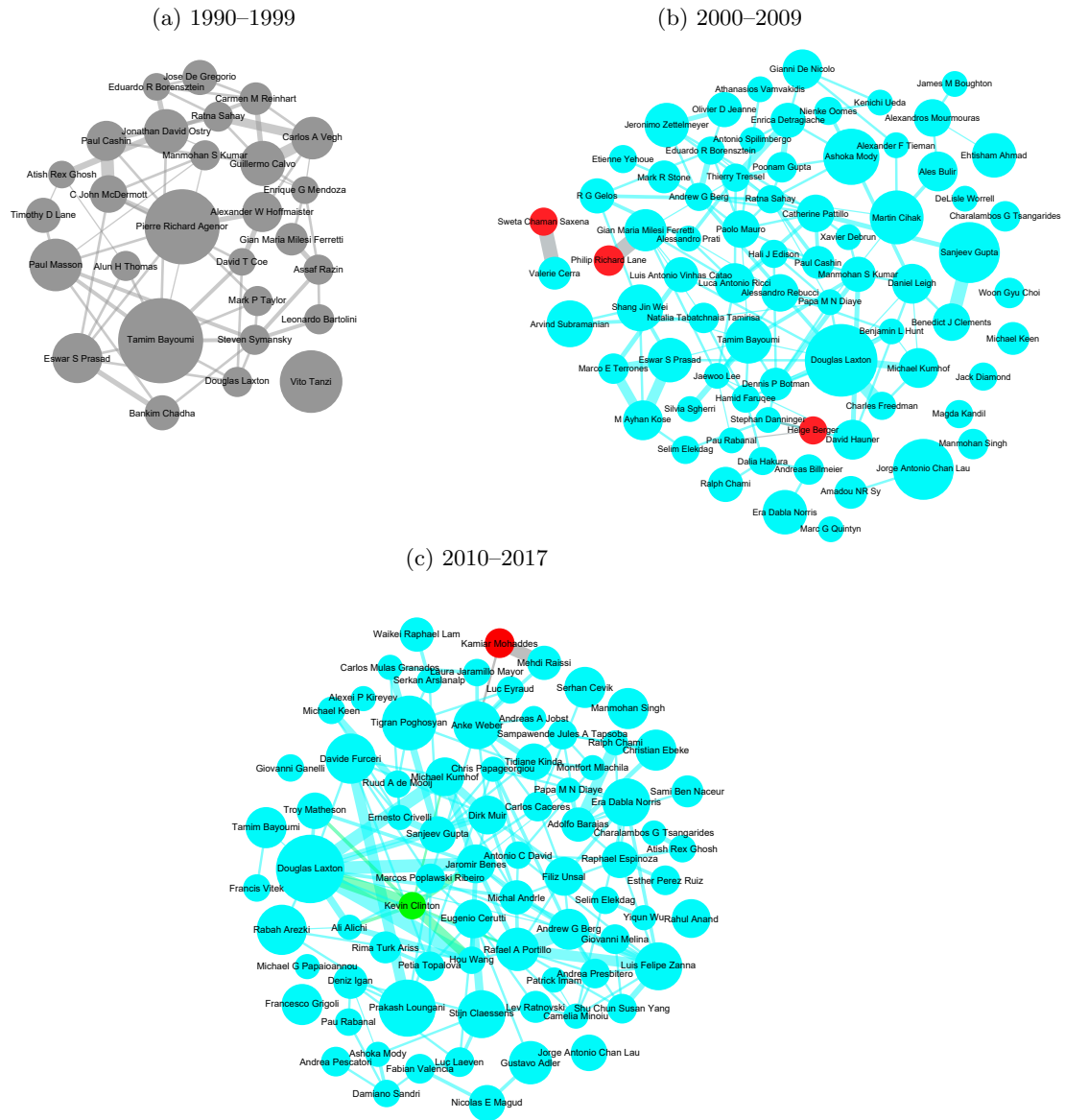
*Notes:* The figure illustrates the distribution of network statistics based on IMF Working Papers co-authorship over 1990–2017.

Figure 16 displays the “core” network over the same three ten-year periods as Figure 14. The network here is limited to authors who contributed to at least ten publications during the respective period. This shows how the set of most prolific authors has evolved over time, with only a few authors appearing in the core network for more than one period.

Another way in which the author-level network can be visualised, and which provides some additional insights into the factors that may drive research collaborations, is by means of clustering techniques. In Figure 17 we take the core network over 2010–2017 (i.e., panel 16c of Figure 16) and subject it to the modularity algorithm of Blondel et al. (2008). This algorithm divides



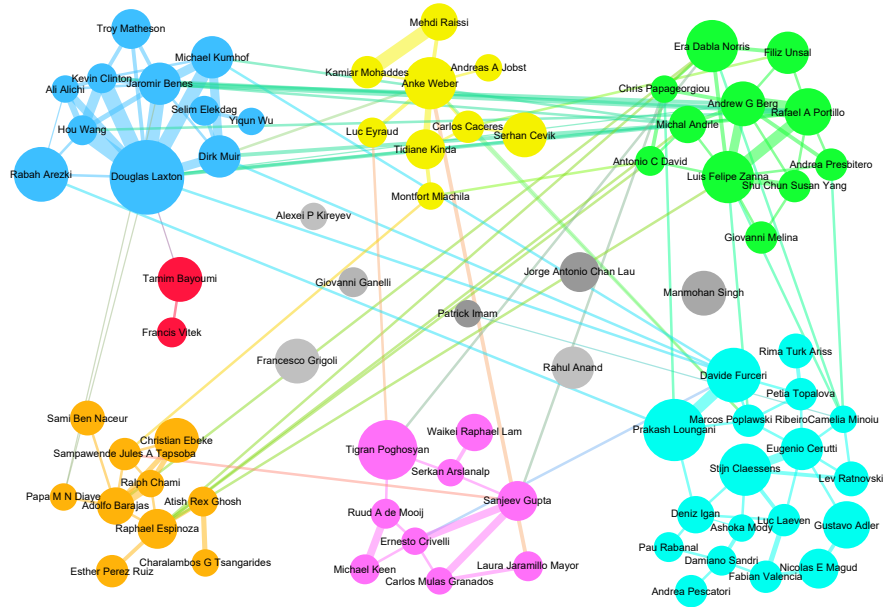
Figure 16: Evolution of 10-year Cumulative Core Author-Level Network



*Notes:* The figure illustrates the evolution of the 10-year cumulative core author-level network based on IMF Working Papers co-authorship over 1990–2017. The sample includes authors who wrote at least ten working papers during the respective period: 1990–1999 (panel 16a), 2000–2009 (panel 16b), and 2010–2017 (panel 16c). The nodes represent authors and edges represent co-authorship. Node size is proportional to the number of working papers written by an author (including single-authored papers) during the respective period. Edge width is proportional to the number of collaborations between two authors during the respective period. In panels 16b and 16c, the node colors are according to the modes of authors' affiliation category over the respective period: light blue denotes the IMF, red denotes universities, light green denotes central banks, dark blue denotes private organizations. Panel 16a discards affiliation categories, as information on affiliations is only available from 1997 onward.

nodes into clusters based on groups of nodes that are more densely connected together than to the rest of the network. The resulting clusters suggest, above all, the importance of working or having worked in the same IMF department as a driver of first-time and subsequent research collaborations at the IMF (in line with Figures 12 and 13). Departmental affiliations tend to be correlated with research areas, another potential collaboration determinant (cf. Figure 6). For example, almost all of the authors included in the top-left sky blue cluster worked at some point in RES and wrote primarily papers on macro-monetary economics with a mathematical component (JEL codes C and E). The green cluster consists of RES and SPR economists. In the pink cluster, we find FAD economists concentrating on research in public economics (JEL code H). And the light blue cluster in the bottom-right corner is mostly RES and WHD staff working on the intersection of international and financial economics (JEL codes F and G).

Figure 17: Clustering of Core Author-Level Network, 2010–2017



*Notes:* The figure illustrates the clustering of the core author-level network based on IMF Working Papers co-authorship over 2010–2017. The sample includes authors who contributed to at least ten publications and is divided into clusters using the modularity algorithm of Blondel et al. (2008). The nodes represent authors and the edges represent publication co-authorship. The node size is proportional to the number of publications by an author (including single-authored papers). The edge width is proportional to the number of publication collaborations between two authors. The node colors identify the clusters to which authors belong. Grey nodes indicate core authors that form a cluster by themselves.

## 4 Determinants of Collaborations

This section investigates the determinants of collaborations across authors. We first describe the empirical strategy. Then, we present the results, along with a number of robustness tests.

### 4.1 Empirical Strategy

It is reasonable to assume that researchers' decisions with whom they collaborate are mostly non-random, based on factors such as complementarity of skills, common research interests, and similarities in background. While some of these dimensions can be easily assessed from publicly available information (e.g., from researchers' publication track record), other relevant knowledge may be only obtained indirectly through the researcher's network of acquaintances, of which the co-author network is an important subset (see Fafchamps et al., 2010 for a formal model). This would imply that, in addition to other pair and individual author characteristics, researchers' positions vis-à-vis each other in the co-authorship network may influence their decision to collaborate.

To study the determinants of first and subsequent collaborations between pairs of authors, we estimate the following logit specification on the sample of active pairs<sup>24</sup>:

$$Pr(y_t^{i,j} = 1) = \Phi(\alpha + \beta N_t^{i,j} + \gamma C_t^{i,j} + \delta HR_t^{i,j} + \epsilon_t^{i,j}) \quad (1)$$

where  $\Phi$  is the logit function and the dependent variable  $y_t^{i,j}$  is, alternatively, the dummy variable tracing first collaborations or the one for subsequent collaborations, as defined in Section 2.1. The set of regressors includes network variables,  $N_t^{i,j}$ ; pairs' characteristics,  $C_t^{i,j}$  (both similar to the ones used in Fafchamps et al., 2010); as well as author-level data on IMF employment and demographics,  $HR_t^{i,j}$ . The variation in our dataset is mostly cross-sectional, as the large majority of author pairs were active only for a short period of time, thereby reducing the need (and feasibility) to control for pair fixed effects.<sup>25</sup>

To compute the network variable regressors, we first define the network graph  $G_t$  with the authors being the nodes and the collaborations between  $t - 9$  and  $t$  being the edges. We allow for a ten-year window as relationships formed at the time of a co-authorship arguably tend to last for some time.<sup>26</sup> The first network variable we include in the regressions is proximity, which is defined as the inverse of the (shortest) distance between the pair of authors,  $d_t^{i,j}$ , in  $G_t$ . When computing  $d_t^{i,j}$ , however, any co-authorship between  $i$  and  $j$  is ignored, so that the minimum distance is equal to two (i.e., the pair has a common co-author) and maximum proximity is equal to 0.5 (even after the initial collaboration), allowing for the use of the variable even as a potential determinant of subsequent collaborations. If there is no link between  $i$  and  $j$ , proximity takes the value of zero.

---

<sup>24</sup>Unlike Fafchamps et al. (2010), we do not draw random samples. Instead we use all pair-year observations for active pairs. In the case of initial collaborations, the sample consists of 6,040,424 observations for 2,310,620 pairs, and in the case of subsequent collaborations the sample consists of 14,893 observations for 3,325 pairs.

<sup>25</sup>While we control for some time-invariant demographic characteristics of the author pair, there might be other time-invariant characteristics that are relevant for the decision to collaborate. Including fixed effects to control for those characteristics, however, would lead to a drastic reduction in the number of observations as there are many author pairs that have been active and collaborated only for one year, and many author pairs that were active for longer but never collaborated.

<sup>26</sup>This leads to the exclusion of the first nine years of the sample from the regressions dataset.

The second network variable is the number of shortest paths between authors  $i$  and  $j$  in  $G_t$ , which is also computed excluding the direct link between  $i$  and  $j$ . If the two authors are not connected at all, the number of shortest paths takes the value zero.

The pairs’ characteristics include research ability, the propensity to collaborate, and research (area) overlap. We proxy research ability with a measure of research productivity. More productive or talented authors are, in general, looked up to and other authors may want to connect with them. As a result, they may be more likely to start new collaborations. At the same time, relatively talented authors may be more interested in working with equally talented co-authors. Hence, we compute a productivity score,  $q_t^i$ , for each author  $i \in G_t$  as the number of “points” author  $i$  accumulated between  $t - 9$  and  $t$ , where points for each published working paper are given by:

$$Points = \frac{citations}{number\ of\ authors + 1}$$

where we add one to the denominator of the ratio to avoid a very high score for papers with few co-authors.<sup>27</sup> We use the average of the productivity scores for each couple of authors to get an aggregate productivity measure for the pair. We also compute the absolute value of the difference in the productivity scores of the two authors to proxy the difference in research ability. The sign of the latter variable will depend on whether more productive researchers look to match with each other or whether co-authorship reflects collaborations between experienced researchers and juniors. In the regression we include the one-year lag of both the average and the absolute difference in productivity to mitigate simultaneity concerns.

A researcher’s propensity to collaborate is also likely to affect the probability of starting new collaborations. A researcher with many collaborators in the recent past may be intrinsically more eager to also collaborate in the future. Moreover, if an author is already well-connected, he is easier to reach out to and this may lead to new collaborations. A proxy for that propensity is the number of co-authors that each author had over the past ten years. The measures of propensity to collaborate for the pair are the lagged average of each author’s propensity to collaborate, as well as the lagged absolute value of the difference in the propensities.

Research overlap is linked to shared research interests and thus likely to condition the probability that two authors start a new collaboration. Researchers sharing similar interests are probably attending similar conferences and events, which creates opportunities to connect and start collaborations. We measure research overlap as in Fafchamps et al. (2010), relying on the JEL codes we discussed in Section 2.2. Specifically, we define the field of research,  $f$ , with the first letter of the JEL classification; and calculate for each author  $i$  the fraction of working papers  $x_{[t-9,t],f}^i$  that has been dedicated to each JEL classification letter between  $t - 9$  and  $t$ . Working papers with multiple JEL codes are assigned proportionally to each of the different codes. We then calculate the research overlap between authors  $i$  and  $j$ ,  $w$ , using the cosine similarity function:

---

<sup>27</sup>Our measure of productivity is somewhat different from the one used by Fafchamps et al. (2010) in that we use the number of citations instead of their “journal quality index” and we do not include the number of pages in the numerator of the ratio. These adjustments reflect that we consider working papers instead of published articles and the general idea that working paper length depends more on the subject than on the quality of the paper.

$$w_t^{i,j} = \frac{\sum_{f \in JEL} (x_{[t-9,t],f}^i x_{[t-9,t],f}^j)}{\sqrt{\sum_{f \in JEL} (x_{[t-9,t],f}^i)^2 \sum_{f \in JEL} (x_{[t-9,t],f}^j)^2}}$$

and we include its lag in the regressions. For subsequent collaborations, the calculation of research overlap excludes the papers on which the authors of the pair previously worked together.

Finally, we also control for other employment and demographic characteristics which appear to matter for collaboration based on the descriptive analysis in Sections 2.2 and 3.2. Specifically, we include a set of dummy variables indicating whether the pair of potential co-authors work in the same IMF department or did so in the past year; whether their gender is the same; and whether their region of citizenship is the same.<sup>28</sup>

## 4.2 Results

We first present the results for the determinants of starting a new collaboration. The sample includes both collaborating author pairs and author pairs that never collaborated (but that were active). In Table 2, we report the logit regression results in terms of odds ratios. Column (1) shows that network distance is a key determinant of first collaborations. For example, for a pair of unconnected researchers (i.e., with proximity at zero) that becomes connected via a common co-author (i.e., with proximity at 0.5), the odds of starting a new collaboration are about 30 times larger; similarly, doubling the number of shortest paths connecting the researchers in the pair increases the probability of initiating a collaboration by a factor of nearly four. These effects are large, which is expected in the case of new collaborations.

In column (2), we include the pairs' characteristics. As these variables enter the specification with a lag, our sample reduces to those author pairs that remained active for at least two years in a row. By excluding "one-off researchers", this sample is likely to be more representative of the relationships across authors that have a somewhat more persistent interest in research. Distance variables have even larger odds ratios, suggesting that more research-oriented people are even more prone to rely on the network around them. The average productivity of the pair of authors is significant only when other variables are included, suggesting that more productive author pairs are more likely to start collaborations. Relying on the significant coefficients in column (3) and (4), we find that a one standard deviation increase in average productivity is associated with a 9 to 10 percent higher probability of starting a new collaboration. At the same time, when the individual productivity of the authors forming the pair diverges, the probability of starting a new collaboration declines, but the effect is not robust to the inclusion of all other determinants. A larger average propensity to collaborate—measured by the number of connections for each author of the pair—is negatively associated with the probability of starting to collaborate, suggesting that authors with many connections (generally more senior staff members with an established network, more options in terms of potential collaborations, and less time to allocate to research) may be less eager to look for new ones. However, if the authors forming the pair have different

<sup>28</sup>This information is available for all researchers that have been affiliated to the IMF at some point during their career.

propensities to collaborate, then they are more likely to start a collaboration. One possible explanation for these results is that senior staff members are often tasked with more managerial duties compared to junior staff, who has more time for research. Therefore, these findings may to some degree reflect an efficient allocation of resources.

In column (3), we include research overlap, which turns out to be insignificant for the probability of starting a new collaboration. This finding is easier to understand when we include the employment and demographic variables in column (4).<sup>29</sup> The odds of starting a new collaboration for author pairs in the same department increases by a factor of more than five, compared to author pairs of researchers from different departments. That is, authors team up because they are in the same department rather than because they dealt with similar research topics in the past. Having citizenship of the the same region also favors the start of new collaborations, by a factor of about 1.3. This may reflect the importance of similar backgrounds, including native languages, and education and professional experiences, among others.<sup>30</sup> Finally, in the presence of the other variables, being of the same gender is not significantly related to the likelihood of starting to work together.

We now turn to the determinants of continuing to collaborate. Table 3 presents the estimation results when the dependent variable is the dummy variable for subsequent collaborations. The number of observations is many times smaller than for initial collaborations, because the dependent variable is defined only for author pairs that collaborated at least once. The results in column (1) suggest that proximity and the number of shortest paths—calculated excluding the direct connections between each pair of researchers—are associated with a higher probability of continuing to collaborate, but the effect is unsurprisingly much smaller than for first collaborations (cf. Fafchamps et al., 2010). For example, reducing the distance between a pair of researchers from three to two almost doubles the chances of continuing to collaborate (compared to an increase by a factor of 11 in the case of initial collaborations); similarly, doubling the number of shortest paths connecting the researchers in the pair increases the probability of initiating a collaboration by about 50 percent. In other words, the network distance between authors is particularly important to start a new collaboration, but once the connection is established, it becomes much less relevant for continuing to collaborate. This result is in line with the interpretation that collaboration networks may transmit important, not directly observable information about authors that helps to overcome matching frictions. Once this information is internalized through actual collaboration, the network effects strongly diminish.

The results for the author pairs' characteristics are very similar to the ones for initial collaborations, as shown in column (2). Specifically, author pairs with a larger average productivity are more likely to continue collaborating. However, if the difference in terms of productivity between the authors of the pair is large, they are less likely to continue working together. The number of established connections, which measures the propensity to collaborate, reduces the probability

---

<sup>29</sup>The inclusion of employment and demographic characteristics purges many external researchers from the sample. However, those that were IMF staff at some point during their career are still part of the sample.

<sup>30</sup>This result is in line with studies such as Freeman and Huang (2015) and AlShebli et al. (2018), who find strong evidence of positive assortative matching on ethnicity in co-authorship networks. Dropping authors that have a European citizenship (excluding transition countries), which represent more than 30 percent of the total number of authors in our sample, does not affect our results.

Table 2: Determinants of First Collaborations

|   | (1)                 | (2)                  | (3)                  | (4)                  |
|---|---------------------|----------------------|----------------------|----------------------|
| Proximity                                   | 6.309***<br>(0.141) | 11.816***<br>(0.578) | 19.365***<br>(1.374) | 20.576***<br>(2.231) |
| Ln number of shortest paths                 | 3.826***<br>(0.149) | 12.052***<br>(0.823) | 19.987***<br>(1.709) | 20.078***<br>(2.349) |
| Lag avg. productivity                       |                     | 1.003<br>(0.002)     | 1.006***<br>(0.002)  | 1.007***<br>(0.002)  |
| Lag abs. diff. in productivity              |                     | 0.996**<br>(0.002)   | 0.997**<br>(0.001)   | 0.998<br>(0.001)     |
| Lag avg. propensity to collaborate          |                     | 0.821***<br>(0.006)  | 0.819***<br>(0.006)  | 0.795***<br>(0.007)  |
| Lag abs. diff. in propensity to collaborate |                     | 1.070***<br>(0.005)  | 1.061***<br>(0.005)  | 1.082***<br>(0.006)  |
| Lag of research overlap                     |                     |                      | 1.030<br>(0.075)     | 0.951<br>(0.083)     |
| Same department                             |                     |                      |                      | 5.096***<br>(0.238)  |
| Same gender                                 |                     |                      |                      | 1.007<br>(0.047)     |
| Same region of citizenship                  |                     |                      |                      | 1.278***<br>(0.067)  |
| Observations                                | 6,040,424           | 3,729,804            | 2,374,994            | 1,921,742            |
| Pairs                                       | 2,310,620           | 875,197              | 682,589              | 526,841              |
| Pseudo $R^2$                                | 0.361               | 0.431                | 0.478                | 0.511                |

*Notes:* The table reports the odds ratios of the logit regressions of first collaborations. Variable definitions are explained in the text. Proximity is multiplied by 10. All regressions include a constant. Heteroskedasticity and autocorrelation robust standard errors in parentheses.

\*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$ .

of continuing to collaborate. And larger differences in terms of the propensity to collaborate increases such probability. As in the case of new collaborations, these results can be explained by referring to the seniority of staff and the allocation of resources.

When we add research overlap in column (3), we find that it has a negative and significant effect on the probability of continuing to collaborate. However, once we include the employment and demographic variables in column (4), research overlap turns insignificant, while being employed in the same department and having citizenship of the same region are still important determinants of the probability of continuing to co-author, increasing the odds by more than a factor three and by 22 percent, respectively.<sup>31</sup> Being of the same gender now leads to an increase in the likelihood of continuing to collaborate of 24 percent.

A commonly used tool to assess the classification performance of logit (and other binary) response models is the receiver operating characteristic (ROC) curve (see Fawcett, 2006). ROC curves visualize the trade-off between the true positive rate—here the ratio of correctly called collaborations by author pairs to the true number of collaborations—and the false positive rate—the ratio of author pairs incorrectly classified as non-collaborating to the true number of non-

<sup>31</sup>As in the case of initial collaborations, dropping authors with a European citizenship (excluding transition countries) does not change the results.

Table 3: Determinants of Subsequent Collaborations

|   | (1)                 | (2)                 | (3)                 | (4)                 |
|---|---------------------|---------------------|---------------------|---------------------|
| Proximity                                   | 1.011<br>(0.022)    | 1.054**<br>(0.026)  | 1.040<br>(0.025)    | 1.094**<br>(0.041)  |
| Ln number of shortest paths                 | 1.507***<br>(0.077) | 2.282***<br>(0.141) | 2.217***<br>(0.138) | 2.508***<br>(0.196) |
| Lag avg. productivity                       |                     | 1.014***<br>(0.002) | 1.016***<br>(0.002) | 1.012***<br>(0.002) |
| Lag abs. diff. in productivity              |                     | 0.992***<br>(0.002) | 0.992***<br>(0.002) | 0.994***<br>(0.002) |
| Lag avg. propensity to collaborate          |                     | 0.900***<br>(0.011) | 0.919***<br>(0.011) | 0.926***<br>(0.011) |
| Lag abs. diff. in propensity to collaborate |                     | 1.067***<br>(0.007) | 1.056***<br>(0.007) | 1.054***<br>(0.008) |
| Lag of research overlap                     |                     |                     | 0.573***<br>(0.063) | 0.898<br>(0.109)    |
| Same department                             |                     |                     |                     | 3.270***<br>(0.253) |
| Same gender                                 |                     |                     |                     | 1.242***<br>(0.104) |
| Same region of citizenship                  |                     |                     |                     | 1.223**<br>(0.098)  |
| Observations                                | 14,893              | 14,589              | 14,207              | 12,645              |
| Pairs                                       | 3,325               | 3,289               | 3,244               | 2,865               |
| Pseudo $R^2$                                | 0.011               | 0.033               | 0.036               | 0.090               |

*Notes:* The table reports the odds ratios of the logit regressions of subsequent collaborations. Variable definitions are explained in the text. Proximity is multiplied by 10. All regressions include a constant. Heteroskedasticity and autocorrelation robust standard errors in parentheses.

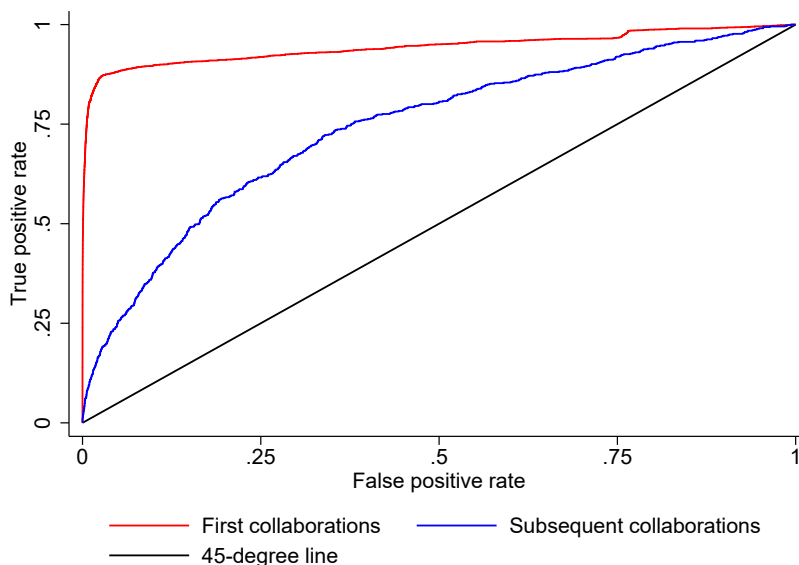
\*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$ .

collaborating pairs—for all possible probability thresholds. The further the ROC curve lies from the 45 degree line, which represents a non-informative classification strategy of random guessing (i.e., the more the ROC curve bends to the top left corner of the plot), the better it is able to discriminate between collaborations and non-collaborations. Figure 18 shows the ROC curves of the richest specifications corresponding to column (4) of Table 2 for first collaborations and column (4) of Table 3 for subsequent collaborations. Both in the case of initial collaborations and subsequent collaborations, the ROC curves are relatively far from the 45-degree line. Specifically, the area under the ROC curve (or AUROC) for the model of initial collaborations is 0.94 and for the model of subsequent collaborations it is 0.74, compared to 0.5 in the case of random guessing. This suggests that with a probability of 94 percent (74 percent) our model for initial (subsequent) collaborations will assign a higher predicted value to a randomly selected collaboration than to a randomly chosen non-collaboration.<sup>32</sup> We conclude that our logit models have a satisfactory performance in terms of separating collaborating from non-collaborating (or no longer collaborating) author pairs.

<sup>32</sup>The relatively weaker performance of the model for subsequent collaborations may be related to the significant reduction in sample size.



Figure 18: ROC Curves



*Notes:* The figure shows the ROC curves of the richest logit model specifications corresponding to column (4) of Table 2 for first collaborations and column (4) of Table 3 for subsequent collaborations.

#### 4.2.1 Robustness

In Table 4, we present a set of robustness tests which rely on modifications of the richest specifications in Tables 2 and 3. First of all, beyond collaborating among themselves, IMF staff collaborate with external researchers (or with researchers that were affiliated to the IMF at some point during their career and are now external researchers). To ensure that our results are not solely driven by the collaborations with external researchers, we exclude them from the sample altogether and re-estimate the logit models.<sup>33</sup> The results in column (1) for initial collaborations and in column (2) for subsequent collaborations are very similar to the baseline.

Second, it could be argued that not all collaborations in the sample are equally “novel”. For example, assume that both author  $i$  and author  $j$  collaborated with author  $h$  in the past, on separate or joint papers. If now all three authors start (or continue) collaborating on a new joint paper, the new link formed (or repeated collaboration) between authors  $i$  and  $j$  may be more related to the presence of middleman/woman  $h$  on the paper project than to their close position in the network per se (the minimum distance of two in this case). That is, without the involvement of author  $h$ , authors  $i$  and  $j$  might never have become (or continued to be) co-authors. A co-authored paper by just authors  $i$  and  $j$ , without middleman/woman  $h$  (but perhaps facilitated by his/her referral of  $i$  to  $j$ ) would arguably constitute a more genuinely new (or repeated) collaboration. To test whether our main results still hold once the “middleman/woman effect” is neutral-

<sup>33</sup>This implies the exclusion of visiting and resident scholars too, since we classified those as external authors affiliated with their home universities (see Section 2.1).

ized, we re-estimate our logit models after excluding from the sample all collaborations of author pairs in papers involving at least one other author that collaborated (either separately or jointly) with both authors of the pair in the past. This leads to the exclusion of 17 author pairs for initial collaborations and eight author pairs for subsequent collaborations. Column (3) and column (4) report the results for initial collaborations and subsequent collaborations, respectively, which are again very close to the baseline.

Finally, instead of defining the network of authors and collaborations over the period  $t - 9$  to  $t$ , we define it over the shorter period  $t - 4$  to  $t$ . This allows us to work with a larger sample, as less data is now required to compute the network variables. The results in column (5) for initial collaborations and in column (6) for subsequent collaborations largely confirm the baseline results. Only the lag of average productivity turns insignificant for initial collaborations.

Table 4: Determinants of First and Subsequent Collaborations—Robustness

|   | Excluding non-IMF employees |                     | Excl. collab. of pairs with common coauthors |                     | Network defined over $[t - 4, t]$ |                     |
|---|-----------------------------|---------------------|--|---------------------|-----------------------------------|---------------------|
|   | (1)                         | (2)                 | (3)  | (4)                 | (5)                               | (6)                 |
| Proximity                                   | 13.454***<br>(1.023)        | 1.113**<br>(0.057)  | 14.247***<br>(1.099)                         | 1.067*<br>(0.037)   | 7.036***<br>(0.338)               | 1.291***<br>(0.041) |
| Ln number of shortest paths                 | 15.214***<br>(1.443)        | 2.818***<br>(0.273) | 15.109***<br>(1.444)                         | 1.863***<br>(0.177) | 8.629***<br>(0.660)               | 2.369***<br>(0.186) |
| Lag avg. productivity                       | 1.003*<br>(0.003)           | 1.016***<br>(0.003) | 1.004*<br>(0.002)                            | 1.015***<br>(0.002) | 1.002<br>(0.003)                  | 1.021***<br>(0.003) |
| Lag abs. diff. in productivity              | 0.999<br>(0.002)            | 0.993***<br>(0.002) | 0.998<br>(0.002)                             | 0.993***<br>(0.002) | 0.999<br>(0.002)                  | 0.990***<br>(0.003) |
| Lag avg. propensity to collaborate          | 0.819***<br>(0.007)         | 0.933***<br>(0.013) | 0.799***<br>(0.007)                          | 0.856***<br>(0.018) | 0.806***<br>(0.008)               | 0.905***<br>(0.014) |
| Lag abs. diff. in propensity to collaborate | 1.072***<br>(0.006)         | 1.048***<br>(0.009) | 1.081***<br>(0.006)                          | 1.075***<br>(0.012) | 1.066***<br>(0.007)               | 1.073***<br>(0.010) |
| Lag of research overlap                     | 0.940<br>(0.083)            | 1.087<br>(0.164)    | 0.898<br>(0.077)                             | 1.078<br>(0.163)    | 1.047<br>(0.083)                  | 0.843<br>(0.104)    |
| Same department                             | 5.653***<br>(0.297)         | 4.626***<br>(0.484) | 5.734***<br>(0.284)                          | 3.159***<br>(0.298) | 5.092***<br>(0.256)               | 2.824***<br>(0.219) |
| Same gender                                 | 1.053<br>(0.050)            | 1.250**<br>(0.124)  | 1.009<br>(0.047)                             | 1.305**<br>(0.138)  | 1.070<br>(0.047)                  | 1.299***<br>(0.108) |
| Same region of citizenship                  | 1.242***<br>(0.067)         | 1.130<br>(0.110)    | 1.282***<br>(0.067)                          | 1.310***<br>(0.126) | 1.300***<br>(0.064)               | 1.265***<br>(0.100) |
| Observations                                | 1,449,272                   | 9,962               | 1,921,631                                    | 12,346              | 2,039,812                         | 13,026              |
| Pairs                                       | 421,217                     | 2,495               | 526,824                                      | 2,857               | 559,031                           | 2,901               |
| Pseudo $R^2$                                | 0.511                       | 0.121               | 0.501  | 0.071               | 0.486                             | 0.117               |

*Notes:* The table reports the odds ratios of the logit regressions of first and subsequent collaborations for a set of robustness tests. Variable definitions are explained in the text. Proximity is multiplied by 10. All regressions include a constant. Heteroskedasticity and autocorrelation robust standard errors in parentheses.

\*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$ .

## 5 Conclusions

In this paper, we study the formation of new and repeated research collaborations using the co-authorship network of the near-universe of IMF Working Papers published during 1990–2017. By not being subject to the “publish-or-perish” conditions of the academia and by being open to external collaborations, the IMF Working Papers series provides an appropriate framework to study the endogenous nature of co-authorship formation.

We document that the number of co-authors per paper has steadily increased over time. The large majority of authors, however, published only one paper during the period under consideration and very few authors wrote more than 20. IMF research authorship has progressively become more diverse reflecting changes in the IMF’s workforce, with women and non-U.S. and non-European authors increasing their shares in total working paper output. We also find that while most working papers deal with macroeconomics, international economics, financial economics, and public economics, in recent years increased attention has been devoted to less conventional topics for the IMF, including climate change and gender.

Across the network—which became larger and more integrated over the three decades of the sample period—IMF staff dominates in terms of the number of authors and working papers, followed at a clear distance by academics, central bankers, and staff from other international organizations. Authors affiliated with government agencies, private companies, and non-profits occupy a much less prominent role in the network. At the institution level, the most recurrent links are with researchers from the World Bank and U.S. and U.K. universities, and in second instance, from other European universities and advanced economy central banks. Among the institutions involved in at least ten papers, there are only a few from emerging market or developing economies and no government agencies or non-profits. Notwithstanding the links with many different institutions, research collaborations within the same IMF department are more common than co-authorship across different departments or with external institutions. Network statistics further indicate that the IMF Working Papers co-authorship network exhibits “small world” properties similar to those of the network of economics journals: it consists of many authors with few direct co-authors, yet indirectly connected to each other through short co-authorship chains, and is characterized by a significant degree of clustering.

The empirical investigation of the determinants of co-authorship relations reveals, above all, that two researchers that are closer to each other in the existing co-authorship network are significantly more likely to collaborate, especially so in the case of first-time collaborations. This result is in line with the interpretation that the network transmits important, not directly observable information about authors that helps to overcome matching frictions, before this information is internalized through actual collaboration. In addition, we find that author pairs with a higher average productivity are more likely to start and continue collaborations. At the same time, researchers with an established network of co-authors seem to be less in search of new ones. Yet, authors with different co-authorship network sizes are more likely to start and maintain a collaboration, possibly reflecting co-authorship between senior staff members (generally tasked with more managerial duties) and junior staff (who have more time for research). Finally, being employed

in the same department and having citizenship of the same region increase the likelihood of both starting and maintaining a collaboration, while greater overlap in broadly defined research areas does not.

## References

- Adams, Jonathan (2012). “The Rise of Research Networks”. *Nature* 490, pp. 335–336.
- Aizenman, Joshua, Hali Edison, Larissa Leony, and Yi Sun (2011). “Evaluating the Quality of IMF Research: A Citation Study”. IEO Background Paper BP/11/01.
- AlShebli, Bedoor K, Talal Rahwan, and Wei Lee Woon (2018). “The Preeminence of Ethnic Diversity in Scientific Collaboration”. *Nature Communications* 9(5163).
- Azoulay, Pierre, Joshua S Graff Zivin, and Jialan Wang (2010). “Superstar Extinction”. *Quarterly Journal of Economics* 125(2), pp. 549–589.
- Blondel, Vincent D, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre (2008). “Fast Unfolding of Communities in Large Networks”. *Journal of Statistical Mechanics: Theory and Experiment* 2008(10).
- Borjas, George J and Kirk B Doran (2015). “Which Peers Matter? The Relative Impacts of Collaborators, Colleagues, and Competitors”. *Review of Economics and Statistics* 97(5), pp. 1104–1117.
- Boudreau, Kevin J, Tom Brady, Ina Ganguli, Patrick Gaule, Eva Guinan, Anthony Hollenberg, and Karim R Lakhani (2017). “A Field Experiment on Search Costs and the Formation of Scientific Collaborations”. *Review of Economics and Statistics* 99(4), pp. 565–576.
- Campos, Raquel, Fernanda Leon, and Ben McQuillin (2018). “Lost in the Storm: The Academic Collaborations That Went Missing in Hurricane Isaac”. *Economic Journal* 128(610), pp. 995–1018.
- Card, David and Stefano DellaVigna (2013). “Nine Facts about Top Journals in Economics”. *Journal of Economic Literature* 51(1), pp. 144–161.
- Catalini, Christian (2018). “Microgeography and the Direction of Inventive Activity”. *Management Science* 64(9), pp. 4348–4364.
- Chai, Sen and Richard B Freeman (2019). “Temporary Colocation and Collaborative Discovery: Who Confers at Conferences”. *Strategic Management Journal* 40(13), pp. 2138–2164.
- Colussi, Tommaso (2018). “Social Ties in Academia: A Friend is a Treasure”. *Review of Economics and Statistics* 100(1), pp. 45–50.
- Ductor, Lorenzo (2015). “Does Co-authorship Lead to Higher Academic Productivity?”. *Oxford Bulletin of Economics and Statistics* 77(3), pp. 385–407.
- Ductor, Lorenzo, Marcel Fafchamps, Sanjeev Goyal, and Marco J van der Leij (2014). “Social Networks and Research Output”. *Review of Economics and Statistics* 96(5), pp. 936–948.
- Ductor, Lorenzo, Sanjeev Goyal, and Anja Prummer (2018). “Gender and Collaboration”. Cambridge Working Papers in Economics 1820.
- Fafchamps, Marcel, Marco J van der Leij, and Sanjeev Goyal (2010). “Matching and Network Effects”. *Journal of the European Economic Association* 8(1), pp. 203–231.
- Fawcett, Tom (2006). “An Introduction to ROC Analysis”. *Pattern Recognition Letters* 27(8), pp. 861–874.
- Freeman, Richard B and Wei Huang (2015). “Collaborating with People Like Me: Ethnic Coauthorship within the United States”. *Journal of Labor Economics* 33(S1), S289–S318.

- Frenken, Koen, Jarno Hoekman, Suzanne Kok, Roderik Ponds, Frank van Oort, and Joep van Vliet (2009). “Death of Distance in Science? A Gravity Approach to Research Collaboration”. *Innovation Networks, Understanding Complex Systems*. Ed. by Andreas Pyka and Andrea Scharnhorst. Berlin/Heidelberg: Springer-Verlag. Chap. 3, pp. 43–57.
- Goyal, Sanjeev, Marco J van der Leij, and José Luis Moraga-González (2006). “Economics: An Emerging Small World”. *Journal of Political Economy* 114(2), pp. 403–412.
- Hamermesh, Daniel S (2013). “Six Decades of Top Economics Publishing: Who and How?” *Journal of Economic Literature* 51(1), pp. 162–172.
- (2018). “Citations in Economics: Measurement, Uses, and Impacts”. *Journal of Economic Literature* 56(1), pp. 115–56.
- Henriksen, Dorte (2016). “The Rise in Co-authorship in the Social Sciences (1980-2013)”. *Scientometrics* 107(2), pp. 455–476.
- Hollis, Aidan (2001). “Co-authorship and the Output of Academic Economists”. *Labour Economics* 8(4), pp. 503–530.
- Hsieh, Chih-Sheng, Michael König, Xiaodong Liu, and Christian Zimmermann (2018). “Superstar Economists: Coauthorship Networks and Research Output”. CEPR Discussion Paper DP13239.
- IEO (2011). “Research at the IMF: Relevance and Utilization”. Evaluation Report.
- Kuld, Lukas and John O’Hagan (2018). “Rise of Multi-authored Papers in Economics: Demise of the ‘Lone Star’ and Why?” *Scientometrics* 114, pp. 1207–1225.
- Liben-Nowell, David and Jon Kleinberg (2007). “The Link-Prediction Problem for Social Networks”. *Journal of the American Society for Information Science and Technology* 58(7), pp. 1019–1031.
- Martín-Martín, Alberto, Enrique Orduna-Malea, Mike Thelwall, and Emilio Delgado López-Cózar (2018). “Google Scholar, Web of Science, and Scopus: A Systematic Comparison of Citations in 252 Subject Categories”. *Journal of Informetrics* 12(4), pp. 1160–1177.
- McDowell, John M and Michael Melvin (1983). “The Determinants of Co-Authorship: An Analysis of the Economics Literature”. *Review of Economics and Statistics* 65(1), pp. 155–160.
- Medoff, Marshall H (2003). “Collaboration and the Quality of Economics Research”. *Labour Economics* 10(5), pp. 597–608.
- Newman, Mark E J (2001). “The Structure of Scientific Collaboration Networks”. *PNAS* 98(2), pp. 404–409.
- Rath, Katharina and Klaus Wohlrabe (2016). “Recent Trends in Co-authorship in Economics: Evidence from RePEc”. *Applied Economics Letters* 23(12), pp. 897–902.
- Ravallion, Martin and Adam Wagstaff (2012). “The World Bank’s Publication Record”. *Review of International Organizations* 7(4), pp. 343–368.
- Sommer, Vera and Klaus Wohlrabe (2017). “Citations, Journal Ranking and Multiple Authorships Reconsidered: Evidence from Almost One Million Articles”. *Applied Economics Letters* 24(11), pp. 809–814.
- Sutter, Matthias and Martin Kocher (2004). “Patterns of Co-authorship among Economics Departments in the USA”. *Applied Economics* 36(4), pp. 327–333.

Wuchty, Stefan, Benjamin F Jones, and Brian Uzzi (2007). “The Increasing Dominance of Teams in Production of Knowledge”. *Science* 316(5827), pp. 1036–1039.