

An Empirical Investigation of the Networks of Scientific
Collaboration in Development Economics: Evidence
from National Bureau of Economic Research's Working
Paper Series

Hanna Lee and Qiwei Lin

December 9, 2021

Contents

1	Introduction	3
2	Research Design	4
2.1	Data	4
2.2	Network Definition	5
3	Result	6
3.1	Co-authorship Network	6
3.1.1	Statistical Properties	6
3.1.2	Prominent Scholars and Scholar Pairs	7
3.1.3	Productivity and the Legitimacy of Researchers	8
3.1.4	Betweenness and Funneling	10

3.1.5	Case study of Edward Miguel’s Ego network	12
3.1.6	Differential returns of social capital by gender	12
3.2	Scholar-Implementer Network	13
4	Discussion	14
	References	16

List of Tables

1	Summary Statistics for the Co-authorship Network	7
2	Top 10 Most Influential Scholars	8

List of Figures

1	Productivity and Scholar’s Characteristics	9
2	The Share of Total Betweenness Centrality Score by Top Scholars	10
3	The Mean Length of Shortest Paths after Removing Top Scholars	11
4	The Ego Network of Edward Miguel	18
5	Scholar-Implementer Partnership Network	19

1 Introduction

Collaboration within academic circles has long been a subject of interest. Examining the network of scientific collaboration in a discipline provides insights into the dynamics of the academic communities and helps understand potential contextual factors that shape the network and behaviors of scholars within it. Research on collaboration has been conducted on various sub-topics: the concept, sources, and identifying patterns of collaboration and main collaborators, and so forth. Previous literature has defined paper co-authorship as research collaboration (Katz and Martin, 1997; Melin and Persson, 1996) and examined the statistical properties of the co-authorship network in natural science disciplines (Newman, 2004, 2001) and public health (Fonseca et al., 2016). The pattern of scientific collaboration in the field of development economics presents us information critical to understanding the production process of evidence that informs policy practices in international development. This network, however, did not receive enough scholarly attention.

Additionally, the literature of scientific collaboration studies has paid little attention to other forms of collaboration. In the field of development economics, a scholar-implementer partnership is also a prevalent form of collaboration as scholars usually partner with implementing organizations in developing countries for the fieldwork of randomized control trials (RCTs). RCTs have received attention from both policy-makers and scholars in the last two decades and become an increasingly common research design in the field of development economics. Researchers adopting this type of research design usually need to partner with organizations in their country of interest to train survey enumerators, collect baseline data, implement the interventions, and conduct follow-up surveys. Investigations of this type of scholarly collaboration yield valuable information about the production of evidence in development research and the process of translating research evidence into policy and program design.

Thus, we investigate the network of scholarly collaboration in development economics in this paper. More specifically, we examine the pattern of collaboration in the co-authorship network and scholar-implementer partnership network. Although collaboration also exists in the form of

commenting and a research assistantship, we decided to focus on co-authorship and research implementation as these two types of collaboration are major factors in determining the overall structure of newly submitted working papers that have not received peer review. For the co-authorship network, we are interested in the following questions: 1) Who are the influential scholars in development economics? 2) Are there any differential returns of social capital depending on gender in the co-authorship network? If so, how does that affect productivity? 3) How the legitimacy of a researcher in his/her co-authorship network affects his/her productivity? 4) What implementing organizations do scholars usually partner with for the fieldwork of RCTs?

The outline of this paper is as follows. In *Research Design*, we provide a definition for networks of interest and detailed information of our data collection process. In *Result*, we describe the empirical results and discuss the substantive meaning of results. In *Discussion*, we review the main conclusion, identify potential limitations and provide suggestions for future research.

2 Research Design

2.1 Data

We use data collected from the National Bureau of Economic Research (NBER) Working Paper Series to construct collaboration networks in development economics. The NBER is a network of nearly 1,600 economists who hold primary appointments at North American colleges and universities and its working paper series distribute new research by NBER affiliates for discussion and comment. This collection of papers does not represent all the scholarly collaboration in development economics but it includes working papers written by many leading development economists.

We expand the data collected in Davies (2021) in the R package `nberwp`. This dataset includes all working papers submitted to NBER from June 1973 to June 2021. For each paper, we are able to identify authors, the month of submission, and the subfield of economics this paper falls into. Although datasets in `nberwp` include the subfield information for each paper, the subfield of Development Economics was not created in the paper series until September 2012. Development

economics papers submitted before this time might be classified as papers in other subfields. This reduces the coverage of the dataset and introduces an artificial boundary to our study. Given that it is time-prohibitive to go through all papers submitted before 2012 to identify potentially misclassified papers, we will narrow the analysis to focus on development economics papers from September 2012 ($N = 1277$).

We impute the gender of economics scholars using the gender package in R developed by Mullen (2021). With this package, we are able to query the gender information using the first name of authors on the API of Genderize, which is based on "user profiles across major social networks". The strength of this imputation is that we are able to quickly obtain relatively reliable results for scholars from English-speaking or other western countries. However, the imputed gender information is less reliable for scholars from South Asia and East Asia.

For the scholar-implementer network, we only examined development economics papers from June 2019 to June 2021 due to time constraints. For each paper, we identify implementing organizations who worked with authors to implement RCTs from the paper.

All existing datasets and additional data we might need to collect are already publicly available either on NBER's website or the researchers' homepage. This availability itself attests to that the consent to use data is given. The research we are conducting also poses no more risk to study subjects than expected in daily life, thus, we do not anticipate significant ethical issues and conclude that no additional measures for human subject protection need to be taken.

2.2 Network Definition

The co-authorship network is a one-mode network in which each node is a scholar who submitted one or more papers of development economics to NBER. A tie exists between two scholars if they ever co-authored one or more papers.

The scholar-implementer partnership network is a two-mode network with two sets of nodes: the scholar set and the implementer set. The scholar set includes scholars who authored development economics papers while the implementer set includes implementing organizations that con-

duct the fieldwork. A tie exists between two nodes if a researcher has a research partnership with an implementing organization.

3 Result

3.1 Co-authorship Network

3.1.1 Statistical Properties

The co-authorship network has 1984 authors (nodes) and 4247 collaboration ties (edges) after removing authors without collaborators. This is a relatively small network compared to other networks previously examined in Newman (2004). A summary of this network's statistical property is presented in Table 1.

On average, an author had no more than two papers over the study period (mean = 1.91, median = 1). A working paper, on average, is written by three authors (mean = 3.014, median = 3). The density of the network is 0.00216, indicating the network is highly sparse. The clustering coefficient shows that there is a moderate level of clustering in the network. The presence of clustering means that there are triangles of researchers in which everyone has collaboration with everyone else. This might be explained by that researchers could have very frequent collaborations with their colleagues in the same institutes or the same research topic. A researcher is able to establish a connection with any other researchers with the support from 7 other researchers as shown by the mean length of shortest paths (6.31). Finally, only 27.8% of the researchers in this network are female scholars, showing that development economics is also a field with gender imbalance.

We look at the components of this network. A component is defined to be a maximal set of nodes in which every node can reach every other by some paths. The number of components and the share of nodes in the largest component can both reveal the connectivity in the network. There are 135 components with varying sizes in this network, indicating that the network is not fully connected. The largest component consists of about 69.56% of the scholars, which is lower than that from the

Table 1: Summary Statistics for the Co-authorship Network

	Mean	Median	Max	Min	SD
Mean Papers per Author	1.92	1	35	1	2.51
Mean Authors per Paper	3.06	3	17	1	1.35
Component Size	14.7	3	1380	2	118.43
Degree (Number of Coauthors)	4.28	3	60	1	4.58
Normalized Betweenness	0.0014	0	0.09	0	0.0056
Eigen Centrality	0.0058	0	1	0	0.041
Density	0.002				
Clustering Coefficient	0.40				
Mean Length of Shortest Paths	6.31				
Proportion of Female Scholars	27.8%				

natural science collaboration network (ranging from 82% to 92%) as shown in Newman (2004). That said, we still see that a large proportion of scholars in development economics is connected with each other through research collaboration. This is the desired property that indicates that most scholars are somewhat connected to the mainstream instead of working in isolation.

3.1.2 Prominent Scholars and Scholar Pairs

Table 2 presents the Top 10 scholars under three different centrality measures: degree centrality, betweenness centrality and eigenvalue centrality. Edward Miguel from the University of California, Berkeley is identified as the most prominent scholars by degree centrality and the second most prominent by betweenness centrality. The eigenvalue centrality ranking gives a very different picture than that from the other two measures.

We also calculate the strength of the collaboration between pairs of scholars using the measure suggested by Newman (2004). Each paper coauthored by a given author pair adds an amount $\frac{1}{n-1}$ to the strength of their collaboration, where n is the total number of authors on the paper. We inversely weight the strength of connection by the size of collaboration because the author divides his/her time between the $n - 1$ other authors with whom he/she works on a paper. The adjusted measure above indicates that Sebastian Galiani and Gustavo Torrens is the pair of authors with the strongest collaboration, followed by two other pairs: 1) Raymond Fisman and Yongxiang Wang; 2) Michael

Table 2: Top 10 Most Influential Scholars

Degree	Betweenness	Eigenvalue
Edward Miguel	Dean Karlan	Abhijit V Banerjee
Abhijit V Banerjee	Edward Miguel	Esther Duflo
Paul J Gertler	Paul J Gertler	Benjamin A Olken
Esther Duflo	Abhijit V Banerjee	Rema Hanna
Sebastian Galiani	Ahmed Mushfiq Mobarak	Arun G Chandrasekhar
Dean Karlan	Francisco Gallego	Emily Breza
Benjamin A Olken	Cristian Pop-Eleches	Sudarno Sumarto
Ahmed Mushfiq Mobarak	Johannes Haushofer	Matthew O Jackson
Arun G Chandrasekhar	Rajeev H Dehejia	Harini Kannan
Johannes Haushofer	Sujata Visaria	Cynthia G Kinnan

Kremer and Christopher M Snyder.

3.1.3 Productivity and the Legitimacy of Researchers

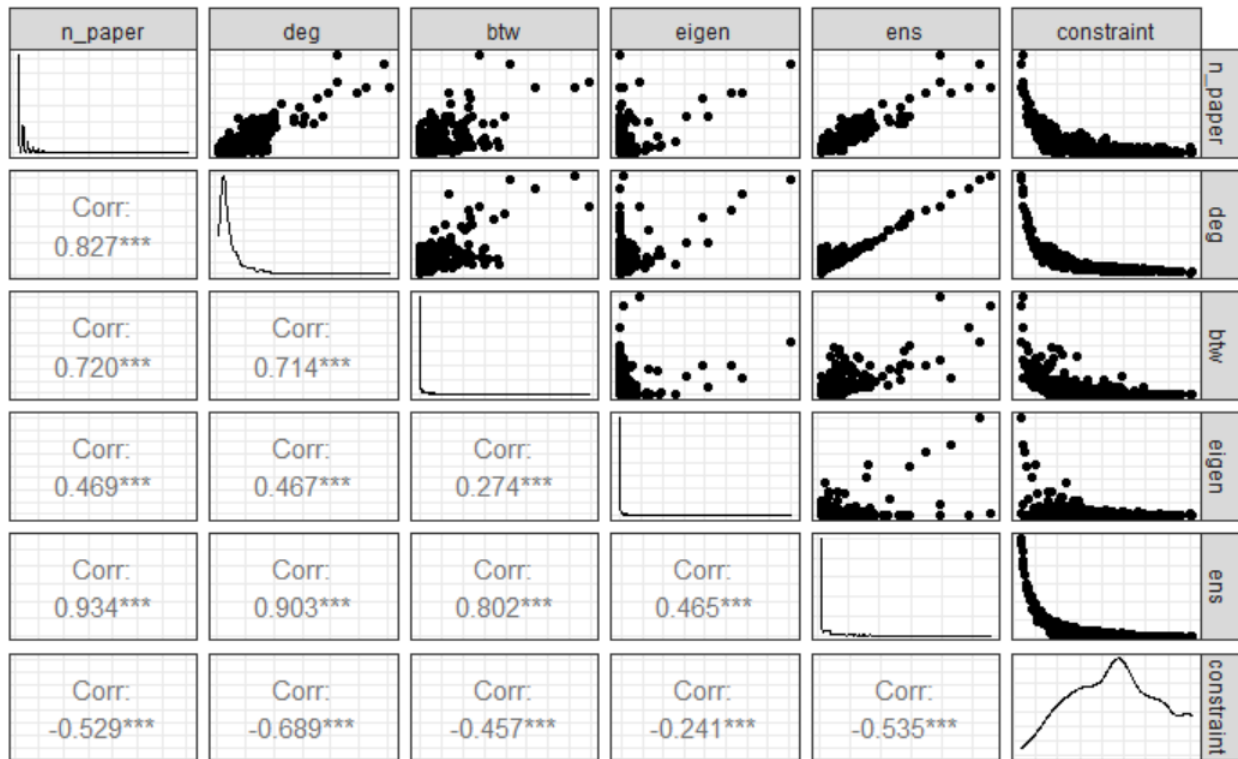
The research question to be addressed in this section is: How the legitimacy of a researcher in his/her co-authorship network affects his/her productivity? We built our research question on the definition of legitimacy by Berger et al. (1998). In their social psychological theory of the legitimization of informal status hierarchies, they define legitimacy as "the process by which cultural accounts from a larger social framework in which a social entity is nested are construed to explain and support the existence of that social entity". This definition informs the research question in a way that a legitimate researcher is understood to be nested in the field of development economics academia, and that legitimate member is supported by the field. Also, in order to tackle this question, we employ the number of papers as the indicator measuring productivity, and constraint score and effective network size by Burt (2004) as indicators measuring legitimacy. Thus, we statistically test the following hypotheses:

- Hypothesis 1: Effective size is positively associated with productivity.
- Hypothesis 2: Constraint is negatively associated with productivity.

The rationale for these hypotheses comes from Burt's Structural Hole Theory. According to the theory (Burt, 2004), nodes with contact networks rich in structural holes—contacts in networks

are scattered, and thus leading to many structural holes in networks—enjoy predicted social capital effect: information and control benefits. The concept of effective size and constraint score are also conceptualized from Burt’s Theory of Structural Hole. Effective size is computed as the ego’s number of alters minus the average degree each alter has to other alters. Thus, a higher value means more critical if removed. Constraint score is a summary measure of the extent to which network alters are connected with each other. Accordingly, a lower value means more critical if removed.

Figure 1: Productivity and Scholar’s Characteristics



Results in Figure 1 is well aligned with this logic: being a broker is associated with higher scholarly productivity. Specifically, a larger effective network size is significantly associated with higher productivity and a greater constraint is significantly associated with lower productivity as we hypothesize. One interpretation is that researchers with lower constraint scores are not as constrained by other researchers. This, in turn, means that a scholar has access to new information outside of a highly connected pool of researchers. In contrast, those with higher constraint values are constrained by his or her circle of well-connected researchers. This means that those constrained

scholars have low access to information outside of what is already circulating in a highly cohesive circle of researchers. Considering the significant association between effective size/constraint score and productivity, we can conclude that being in a broker position confers researchers a competitive edge in being productive by allowing them to access new information outside cliques.

Figure 2: The Share of Total Betweenness Centrality Score by Top Scholars

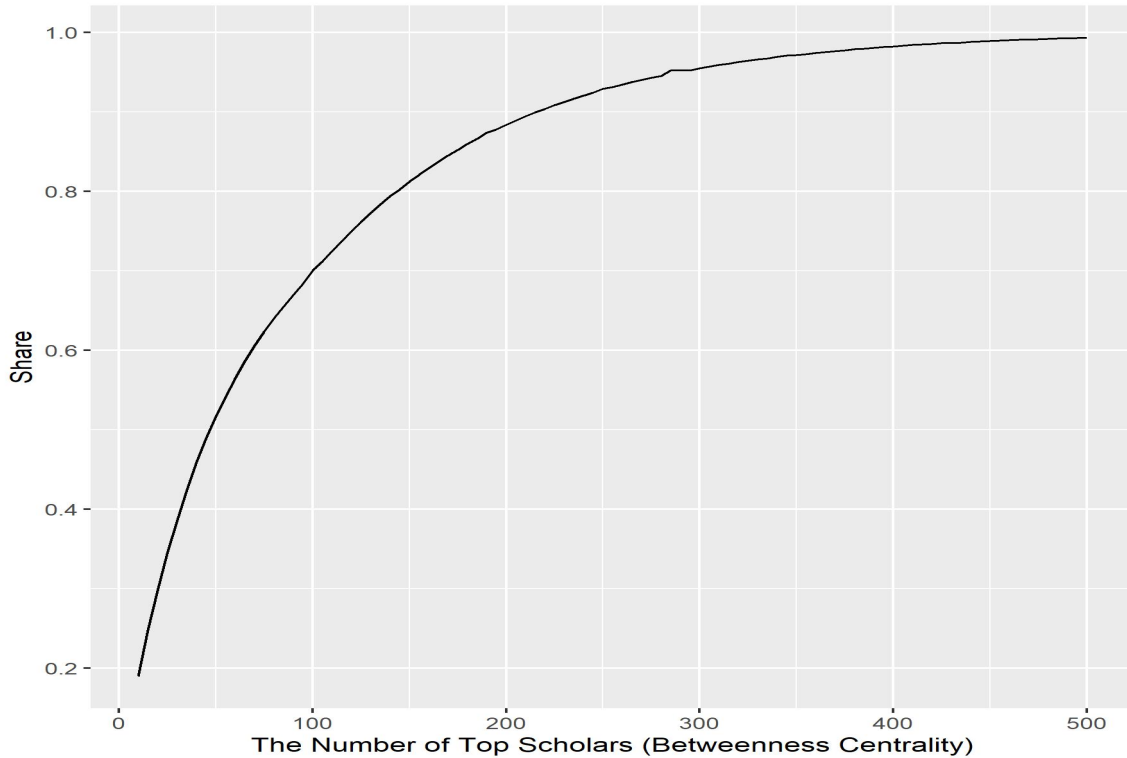
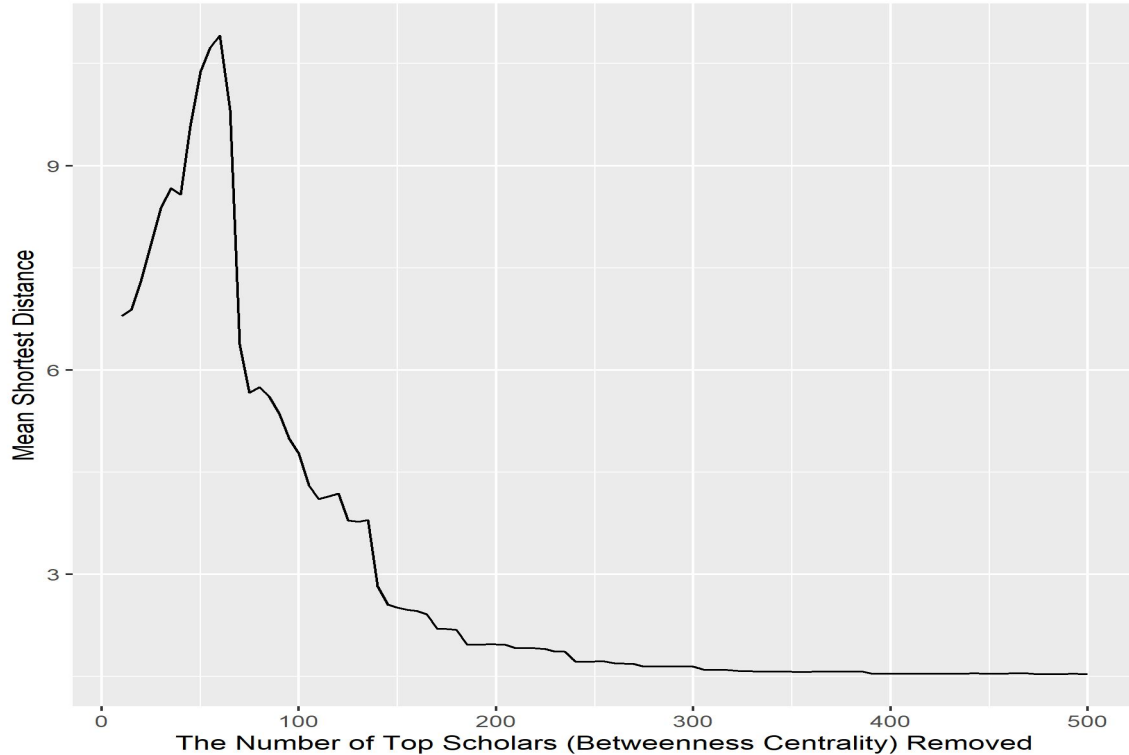


Figure 2 also shows the importance of being in a broker’s position in a co-authorship network. A small number of top scholars have a large share of total betweenness centrality in this co-authorship network.

3.1.4 Betweenness and Funneling

Figure 3 shows that for most authors the bulk of the paths between them and other researchers in the network go through a relatively small number of collaborators. This phenomenon is called funneling according to Newman (2001). More importantly, Figure 3 shows removing the nodes of high betweenness centrality results in large increase in the average distance between other re-

Figure 3: The Mean Length of Shortest Paths after Removing Top Scholars



searchers when they. That is, the current average distance of the whole co-authorship network is 6.31. When we remove the top 60 scholars in terms of betweenness measure, then the average distance sharply increases to 10.91.

This allows us to answer the question that has been raised by Newman (2001): Are all of your collaborators equally important for your connection to the rest of the co-authorship network, or do most connections from other researchers to you pass through just a few of your well-connected collaborators? If Burt's Structural hole theory holds, we expect that the latter would be the case. That is, collaboration with just one or two famous, or "legitimate" members of the development economics's field would easily establish short paths to a large portion of the collaboration network, and all of those short paths would go through those one or two legitimate members.

Our co-authorship network shows that the latter holds. We can expect that for most researchers, also their collaborators would be only a few nodes and most of the paths would be connected between themselves and the rest of the network. The rest of their collaborators, even though they

take up a large portion in the mere number of authors in this network, would take up only a small portion of the total paths.

3.1.5 Case study of Edward Miguel's Ego network

The funneling effect is confirmed with the evidence of betweenness centrality measure as well. 'Betweenness' measures the extent that a node sits 'between' pairs of other nodes in the network, such that a path between the other nodes has to go through that node (Luke, 2015). Thus, the high betweenness of a node indicates that the node is, geodesically speaking, between many researchers. This makes that node prominent because that node is in a position capable of controlling the flow of information between other researchers in the network. This is the case in our co-authorship network. A node that shows the second highest value of betweenness centrality and highest degree is Miguel. Thus, we will take Miguel's ego network to further explore the benefits of being a broker in a co-authorship network as a case study.

In Figure 4, we can see the ties from Edward Miguel connecting researchers who would otherwise have been disconnected. These ties allow Edward Miguel to reach and tap different sources of information in a co-authorship network. In fact, if we remove him from this ego network, this network breaks down to eight different components, indicating that Edward Miguel is connecting eight different alters. In other words, he is capable of both accessing eight different information resources and controlling the flow of the communication among eight different groups: the value of being in a brokerage position. Just as we hypothesize based on Burt's structural hole theory, we expect a node with high betweenness also shows great productivity due to the benefits that node enjoys as a bridge in a network, and it is indeed evident in Edward Miguel's case, who is a node with both high degree and betweenness centrality.

3.1.6 Differential returns of social capital by gender

In this section, we answer the question, "Are there differential returns of social capital depending on gender in co-authorship network? If so, how does that affect productivity?" We initially

hypothesize that there would be differential returns of social capital to different genders, affecting the level of productivity. This conjecture is built on Burt's work, "The Gender of Social Capital". In the article, Burt argues that evidence on women contradicts the predicted social capital effect he stated in the Structural Hole Theory (Burt, 1998). He states that an empirical case in a specific business shows that the entrepreneurial networks associated with success for senior men did not work for women, and thus concludes that women do better with a small network of interconnected contacts unlike their male peers. As this was due to that women were not accepted as legitimate members of a population, we first observed the share of female nodes in our network to see if this case would hold in our data as well. In fact, the share of women researchers in this network is only 27.8%. Thus, with Burt's theory on differential social capital for different genders combined with this figure, we hypothesized the following:

- Hypothesis 1: Constraint is positively associated with productivity for female researchers.
- Hypothesis 2: Constraint is negatively associated with productivity for male researchers

Indeed, just looking at this small share of female researchers in this co-authorship network might give an impression that women are the "illegitimate" members of this network, and thus they are disadvantageous in being productive. However, the correlation between the score of constraint and productivity on different gender groups suggests that the result is opposed to our hypothesis. That is, it turned out that the correlations between constraint and productivity on different sex groups were not different. The correlations are statistically significant and negative for both groups. Thus, we conclude that the correlation between constraint and productivity holds for both genders and thus that there are no differential returns of social capital depending on gender.

3.2 Scholar-Implementer Network

The scholar-implementer network has 100 research institutes, 77 implementing organizations, and 354 collaboration ties. Fig presents the visualization of the largest component in this network

with several well-connected organizations labeled.

Scholars from MIT have a partnership with 25 different NGOs or foreign governments in the developing world to conduct research that utilizes RCTs, and it thus possesses the largest network of research collaboration with implementers. This is not surprising given that two Nobel laureates Abhijit V Banerjee and Esther Duflo, who won the Nobel prize for their experimental approaches to poverty alleviation, are long-time faculty members of economics at MIT. On the implementer side, Innovation for Poverty Action (IPA) has the largest collaboration network. As an organization dedicated to promoting the use of evidence in policy-making, it has partnered with scholars from 62 different research institutes to provide supports on building connections with other organizations, managing research projects and fieldwork. Our result reveals the prominence of MIT and IPA in the field of experimental development economics as expected.

4 Discussion

In this paper, we examine and discuss the structure of collaboration networks in the field of development economics. Top scholars and scholar-pairs are identified and we examined in detail the ego network of Edward Miguel, the scholar with the highest degree centrality and the second-highest betweenness centrality. Our result on the alleged differential returns of social capital by gender contradicts some of the previous literature, particularly Burt on differential social capital for different gender(1998). He stated that evidence on women contradicts the predicted social capital effect of Structural Hole Theory. However, our study suggests that there is no systematic differences in the returns of social capital by gender in co-authorship network in the field of development economics. Finally, the outcome on the association of legitimacy and productivity shed light on strategies researchers can take in co-authorship network: productivity stems from low constraint linked to being rich in structural holes. That is, with a better understanding of strategies associated with higher productivity demonstrated, scholars would be able to make a more informed choice of being in a broker position in a co-authorship network.

Our work is not without limitations. First, we cannot rule out potential aggregation error in co-authorship network (Newman, 2018). That is, for two scholars with the same name, their record of scholarly work could be mistakenly combined into one and thus they might appear as a single node in our co-authorship network. Another potential limitation comes from the coverage of NBER data. This dataset only has development economics paper from June 2012 and does not include informal collaboration among researchers such as informal commenting by colleagues, formal comments by reviewers and editors, and research assistantship. Moreover, we only use the number of submitted working papers as a proxy for productivity. This single indicator might not fully represent the latent variable productivity.

For future research, researchers could further examine nodes identified to be the most productive or prominent researchers and their careers. Doing this with a particular interest in how early a co-authorship network emerge in a form rich in structural holes would give insights into how their network structures evolved. Researchers could garner information on authors' affiliation and their professional titles at different times. Such information would be useful to explore how individual scholar's network has been formed and if the strategies previously demonstrated to be effective hold. Furthermore, future research can extend the analysis to other subfields of economics or economics as a whole by using a more comprehensive collection of scholarly work than NBER working paper series.

References

- Berger, Joseph, Cecilia L. Ridgeway, M. Hamit Fisek and Robert Z. Norman. 1998. “The Legitimation and Delegitimation of Power and Prestige Orders.” *American Sociological Review* 63(3):379–405.
URL: <http://www.jstor.org/stable/2657555>
- Burt, Ronald S. 1998. “The Gender of Social Capital.” *Rationality and Society* 10(1):5–46.
URL: <https://doi.org/10.1177/104346398010001001>
- Burt, Ronald S. 2004. “Structural Holes and Good Ideas.” *American Journal of Sociology* 110(2):349–399.
- Davies, Benjamin. 2021. *nberwp: Catalogue of NBER working papers published between June 1973 and June 2021*.
- Fonseca, B.d., R.B. Sampaio, M.V.d Fonseca and et al. 2016. “Co-authorship network analysis in health research: method and potential use.” *Health Research Policy and Systems* 14.
- Katz, J Sylvan and Ben R Martin. 1997. “What is research collaboration?” *Research policy* 26(1):1–18.
- Luke, Douglas A. 2015. *A user’s guide to network analysis in R*. Springer.
- Melin, Göran and Olle Persson. 1996. “Studying research collaboration using co-authorships.” *Scientometrics* 36(3):363–377.
- Mullen, Lincoln. 2021. *gender: Predict Gender from Names Using Historical Data*. R package version 0.6.0.
URL: <https://github.com/lmullen/gender>

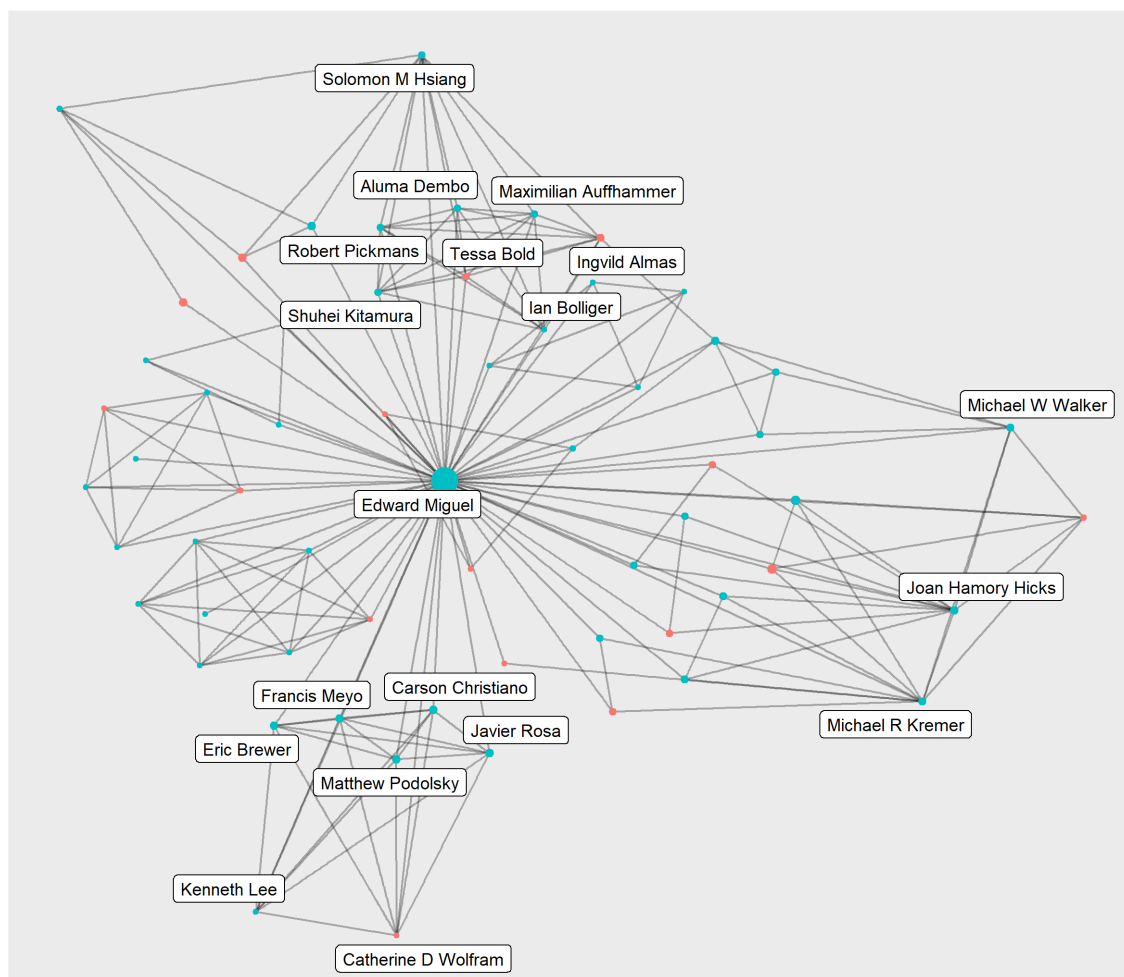
Newman, M. E. J. 2001. “Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality.” *Physics Review. E* 64:016132.

URL: <https://link.aps.org/doi/10.1103/PhysRevE.64.016132>

Newman, M. E. J. 2004. “Coauthorship networks and patterns of scientific collaboration.” *Proceedings of the National Academy of Sciences* 101(suppl 1):5200–5205.

Newman, Mark. 2018. *Networks*. Oxford university press.

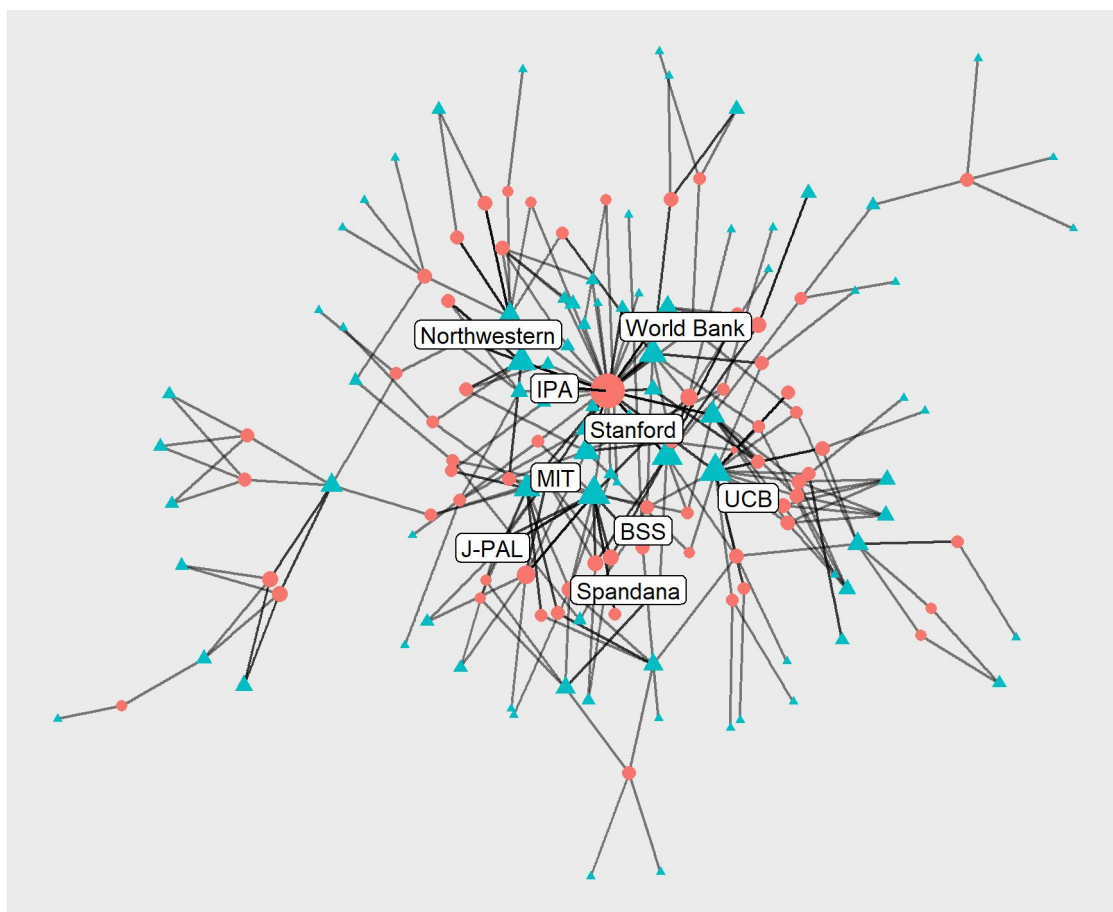
Figure 4: The Ego Network of Edward Miguel



Betweenness Centrality • 0.0 ● 0.2 ● 0.4 ● 0.6 ● 0.8

Male • 0 • 1

Figure 5: Scholar-Implementer Partnership Network



Degree ● 20 ● 40 ● 60

type ● Implementer ▲ Research Agency