Networks in the market for researchers*

Flavio Hafner†

Christoph Hedtrich‡

December 6, 2023

Abstract

We study the role of collaboration networks in the labor market for young scientists in the United States. Nearly one in five PhD graduates publish after graduation at universities where their advisor has a co-author; such pre-existing connections more than double the probability of matching with the university—even within fine-grained peer groups. The importance of the advisor's network for placement doubled from 1990 to 2014. For graduates placed via the advisor's network, we document a citation premium of 9 to 30 percent but find no evidence of private information being revealed.

JEL Codes: J24, J63, M51

*We thank Francis Kramarz, Oskar Nordström-Skans, Bas Hofstra and seminar participants for their valuable comments. We thank Mona Dapfer and Quan (Jack) Cheng Xie for their excellent research assistance. We acknowledge financial support by Jan Wallanders och Tom Hedelius Stiftelse/Tore Browaldhs Stiftelse through program grant P21-0134.

†Netherlands eScience Center, `f.hafner@esciencecenter.nl`

‡Uppsala University, `christoph.hedtrich@nek.uu.se`

Workers and firms heavily rely on social connections for finding matches (Topa, 2011). One possible reason for this is that networks solve an information problem (Rees, 1966): Because neither the firm nor the job candidate can observe all relevant aspects of the match, networks transmit information that lowers search costs. But whether referral hiring leads to better matches is ambiguous. In the optimistic view, referrers promote candidates that are good fits for the job, benefitting firms and workers by increasing their options and match productivity. In the pessimistic view, referrers exploit their relationship with the employer to promote underperforming candidates, leading to nepotism and inefficient matches.

In the market for scientists, PhD advisors can act as referrers, leveraging private knowledge about graduates and connections to potential hiring institutions. Moreover, the growth of research teams and collaboration networks (Fortunato et al., 2018; Wuchty, Jones and Uzzi, 2007; Jones, 2009; Freeman, Ganguli and Murciano-Goroff, 2014) implies a rising role of co-authorship connections—as documented in Figure 1a: The fraction of PhD students that publish their first paper after graduating at a university where their advisor had a pre-existing co-author connection increased from about 10 percent in 1990 to around 25 percent in 2014.

Overall, the rising use of PhD advisors' networks in student placements can impact graduates' careers and science's aggregate productivity. It is therefore important to understand how the collaboration network is used. In this study, we answer the following questions: First, what is the role played by advisors' collaboration networks in placements and productivity outcomes of PhD graduates? Second, do advisors' connections reveal private information about either candidate or match quality at the time of hiring? Third, is the use of adivsors' connections in hiring associated with higher productivity overall? Fourth, has the role of advisors' collaboration networks for the placement of students changed over time?

Our paper is the first to document the link between the rise of scientific collaborations and the labor market for PhD graduates. Moreover, we contribute to the literature on social networks in the labor market more generally by overcoming significant data

constraints in existing work: Existing studies either rely on personnel records from a small number of hiring firms—for example, Brown, Setren and Topa (2016), Burks et al. (2015), and Pallais and Glassberg Sands (2016)—or on population-scale registry data which lack not only direct productivity measures (Eliason et al., 2022) but also direct measures of network connections for a large sample (Dustmann et al., 2016).

Our data have two advantages. First, we closely approximate each graduate's set of potential hiring universities. We can therefore account for any bilateral determinants of hiring flows between PhD-awarding and hiring universities, as well as for the fit in terms of research topics between graduates and potential employers. Second, we provide new insights into the productivity of network hires by comparing graduates with very similar backgrounds and networks. Therefore, we can account for selection into network connections, which otherwise confounds the comparison of network and non-network hires. Third, we show that publicly observable pre-hiring productivity predictors explain differential research output of network and non-network hires.

We start our analysis by showing that advisors' pre-existing collaboration networks significantly impact where their students find their first jobs. We find that having a connection to a university through their advisor's network doubles the probability that the student matches with that university. These results are precisely estimated and robust across specifications where we successively compare employment options across students from the same class, compare employment options of the same student, and account for how similar the student's research topics are to those of the potential hiring university. Looking at time trends, we find a modest decline in the effect of *one* connection on placement. Nevertheless, as the average student in 2014 was connected to three times as many universities as in 1990, the overall importance of advisors' collaboration networks for placing students has doubled.

Our results are not only in line with existing evidence on referrals and networks in the broader labor market (Kramarz and Skans, 2014), but they are also relevant for understanding the academic labor market in particular: We provide evidence for the notion that PhD advisors play a crucial role in matching students to their first job (Long

and McGinnis, 1985). Our evidence also relates to work focusing on hiring networks between universities (Terviö, 2011; Clauset, Arbesman and Larremore, 2015; Anderson and Richards-Shubik, 2021). We document an underlying channel for imbalanced hiring flows between universities in the aggregate.

Our estimates of the effect of advisors' networks on matching have a causal interpretation under the assumption that no factors simultaneously determine where the advisor has a network connection and where the graduate finds their first job relative to peers from the same graduating class. Importantly, these factors need to be unobserved to the analyst but observed by the student on the job market *even without an actual network connection*. Notably, much of the existing work on referrals relies on similar or stronger assumptions (Burks et al., 2015; Brown, Setren and Topa, 2016; Kramarz and Thesmar, 2013).[1]

We then document that graduates hired through their advisor's network produce more output than non-connected hires post-PhD. This aligns with existing work with output data—Pallais and Glassberg Sands (2016) and the high-skill setting in Burks et al. (2015). By delving deeper into the graduates' research output data, we disentangle different sources for this productivity gap. First, we study the gap both from the point of view of the graduating class and from the point of view of the hiring university. Second, we document how the output gap is related to the graduate's and the referrer's pre-match productivity.

When we compare graduates within the same class, we find that those with a pre-existing connection to their new university receive 30% more citations than those with a non-connected first placement. Publicly observable controls—fine-grained fixed effects for the student's pre-graduation impact-weighted output and the advisor's citations—reduce the gap by about half, but it remains statistically significant. Conditional on these controls, connected graduates write more papers, collaborate more with their new colleagues, and are placed at more prestigious universities. In the online appendix, we show very similar results with advisor fixed effects, indicating that the same advisors

[1] An exception is Rajkumar et al. (2022) who conduct an experiment on LinkedIn.

place students through the network that later turn out to be particularly productive.

When we compare new hires at the same university, we find that those with a pre-existing connection produce 14% more citations than those without a connection. The student's pre-graduation productivity alone explains about half of the gap; the other half is explained by including the advisor's citations in the regression. Conditional on these controls, we also find no or minor differences between connected and non-connected hires for other outcomes such as the number of papers and collaborating with colleagues. Thus, the hiring university benefits from access to more productive candidates, but there is no evidence that the connection reveals private information about match productivity.

Finally, our results relate to theories and the broader empirical literature on networks in the labor market. First, the results do not align with models where referrals lead to higher average match productivity through the revelation of private information (Dustmann et al., 2016; Simon and Warner, 1992; Galenianos, 2013). Second, since we find that the productivity of the advisor can explain part of the productivity premium of network hires, our results relate to homophily-based models where firms solicit referrals from productive workers because they are connected to other productive workers (Montgomery, 1991). In the general labor market, this allows firms to hire more productive workers on dimensions that are otherwise unobservable to the firm (Hensvik and Skans, 2016). In our setting, however, the advisor's productivity is publicly observable to all firms, independently of the network connection. Thus, the results indicate that the network does not reveal private information about students' productivity but instead lowers hiring costs within the network, enabling universities to attract graduates who are expected to perform better.

Data

To track publication outcomes, affiliations, and co-author networks of PhD graduates and their advisors, we combine data on PhD dissertations with a large bibliographic

database.[2] The first data set is ProQuest Dissertations & Theses Global (PQDT)™, which has information on PhD dissertations (Proquest, 2021). This information includes the name of the PhD graduate, the name of the advisor(s), the PhD granting university, and the title and abstract of the thesis. The second data set is the Microsoft Academic Graph (MAG), which has information on papers, authors, their affiliations, and citation links (Sinha et al., 2015; Wang et al., 2019).

We link PhD graduates and their advisors with the fuzzy matching algorithm `dedupe` (Gregg and Eder, 2022). Links are identified when two records in the two data sets have similar features. For graduates, this means a similar name, keywords, paper titles, as well as starting their publication career around the year of PhD graduation. For advisors, this means a similar name, affiliation name, and publishing activity around the student's graduation year in the two data sets. `dedupe` works with active learning: the user labels not a random sample of records as true or false links but those potential links that the algorithm is least certain about.

In the online appendix, we discuss the data processing and the quality of the data in more detail. First, both data sources are of high quality. The US Library of Congress uses ProQuest; MAG's coverage of scientific works is comparable to data sources such as Scopus and Web of Science and has been used in previous studies (Huang et al., 2020). Second, while the active learning nature of our linking algorithm prevents us from calculating precision and recall that are representative of the linked sample, we validate the links in two ways. For graduates, we suggest a lower bound on the precision of the linking of 0.78. For advisors, we calculate a recall of about 0.75 across fields of study and years. For details, see the Online Appendix.

Our sample consists of PhD graduates from US universities who received their degrees between 1990 and 2014 and whom we linked to an author in MAG.

We restrict the sample to graduates we observe as (i) publishing at least once within the first seven years after graduating and (ii) doing so with an affiliation to a US university that is not their PhD-granting university. Additionally, we limit our analysis to graduates

[2] Hadlock and Pierce (2021) and Rose and Shekhar (2023) use similar data from different sources to study hiring networks in economics.

for whom we have linked one advisor to MAG.

For each graduate, we construct their connections to universities through their past collaboration network and through the network of their advisor. We also compute vectors of research concepts in graduates' dissertations and by the faculty at potential hiring universities. Concepts are predicted based on abstract text in the dissertation and published papers with the language model provided by Wang et al. (2020, section 2.5), yielding a consistent definition of research concepts between graduates and hiring universities.

Table 1: Summary statistics

Variable	Baseline	Advisor in MAG	Connected first affiliation	Not Connected first affiliation
Degree Year	2004	2004	2006	2004
Year First pub post PhD	2.90	2.88	2.55	2.97
Class Size	21.17	21.71	24.93	20.80
Link Score Student	0.97	0.97	0.97	0.97
Link Score Advisor		0.98	0.97	0.98
Connected Advisor		0.20	0.92	0
Connected Own Co-author		0.05	0.22	0
N Affiliation Connections				
through advisor		12.32	22.22	9.59
through co-author		2.16	4.14	1.62
Outcomes				
Cites 7y post PhD	362	358	581	297
Papers 7y post PhD	11.30	11.34	14.41	10.49
Co-authors first affil	3.76	3.73	6.18	3.05
First Affil 6y post PhD	0.59	0.59	0.48	0.62
Active 6y post PhD	0.89	0.89	0.88	0.89
Cites first affil	7,970	7,805	19,062	4,708
Observations				
N PhD Graduates	88,721	73,885	15,940	57,945
N PhD Graduates × potential hiring institutions	33,639,234	28,477,541	6,097,485	22,380,056

Notes: The baseline sample is PhD graduates in the Proquest Dissertations&Theses database from US universities included in the Carnegie Classification between 1990 and 2014 in all major fields apart from medicine and arts. The column *Advisor in MAG* refers to PhD graduates for whom we have advisor information and have found a link to the Microsoft Academic Graph for the advisor. The rows *Link Score Student* and *Link Score Advisor* are the confidence scores about the identified links between entities. The scores are predicted values from a logit model that compares candidate pairs within blocks of similar records. See the data appendix for details.

Table 1 shows summary statistics for the sample of linked PhD graduates. Each column shows the statistics for a different sample: the first column is for the baseline sample of PhD graduates linked to MAG; the second column is for those with a linked advisor—our primary analysis sample. Columns 3 and 4 split this sample into those with a network connection to their first affiliation and those without.

The baseline sample of graduates linked to MAG consists of 88,721 graduates, for 73,885 of whom we have advisor information and could link the advisor to MAG. This sample is our primary analysis sample.

The table shows that the linking algorithm we use to match supervisors and graduates between PQDT and MAG is very certain about predicted links in all subsamples: On average, the identified links have a score of 97% or more, and this is the same for students and advisors independently of connection status.

In the primary sample, 20% of graduates have a network connection to their first post-PhD affiliation through their advisor, and 5% through their own co-author network. This indicates that network connections can be important for matching PhD graduates to their first post-PhD affiliation, resonating with evidence reported by Brown (1965). We will show to what extent the collaboration network matters for placement and to what extent the rising share of network hires is driven by the growth of collaboration networks or by changes in the importance of the network for matching.

Further, connected graduates are, on average, more productive than non-connected graduates; this holds in terms of the number of papers, number of citations, or year to first publication post-PhD. In the empirical section, we will return to this pattern, where we control for a range of confounders and discuss possible explanations.

Empirical Framework

To show how collaboration networks impact the allocation of PhD graduates to their first job and assess the implications for post-PhD outcomes, we ask two questions. First, to what extent do network connections explain the matching of PhD graduates to their first

university after graduating? Second, do network hires differ in their research output and collaboration patterns at their first university, compared to non-networked hires? We tackle these questions with the following empirical framework.

Network connections and the first affiliation of PhD graduates To assess the empirical relevance of network connections for the transition of PhD graduates to their first post-PhD affiliation, we use the following regression framework (Eliason et al., 2022):

$$\underbrace{E_{i,j}}_{\text{P i matches j}} = \underbrace{\alpha_{c(i),j}}_{\text{class-destination effect}} + \underbrace{\beta_{i,j}X_{i,j}}_{\text{individual controls}} + \underbrace{\gamma A_{i,j}}_{\text{network effect}} + \varepsilon_{i,j}, \tag{1}$$

where each observation is a pair of a graduate i and potential first university j. The variable $A_{i,j}$ indicates whether graduate i has a network connection to university j. We are interested in the parameter γ. It measures the effect of a network connection on the probability of a match. To move towards a causal interpretation of this parameter, we control as finely as possible for various other factors that could affect both the probability of a match and the probability of having a network connection.

First, the fixed effects $\alpha_{c(i),j}$ account for any factors that systematically impact hiring flows between pairs of potential hiring universities and the graduating class. A class is a group of students that graduate in a five-year window from the same university in the same major field—for instance, all Harvard biologists from 2010 to 2014. For example, the fixed effects account for the possibility that universities that are closer to each other—either geographically or in terms of research fields—have more co-authorship connections and are more likely to hire graduates from each other.

Second, we add controls $X_{i,j}$ that vary between the individual student and each potential hiring university. The controls address the concern that matching graduates to universities can depend upon the characteristics of individual graduates, the destination university, and their interaction. A particular concern is specialization within major research fields: An advisor who specializes in molecular biology may have more connections to other biology departments that specialize in molecular biology; if her students also work on molecular biology, they may be more attractive hires for other

molecular biology-specialized departments compared to a student from the same class who specializes in plant biology.

Specifically, we use the cosine similarity between the topic vectors of the student's dissertation and the papers written—in the five years before the student's graduation—at the potential destination university. We calculate this similarity for the collection of all papers and for the papers of the most similar potential collaborator. Our controls thus capture how similar the graduate's research is to the research at the potential employer.

It is also important to discuss which confounders are not captured by our controls and which could invalidate a causal interpretation of the network parameter γ. Such confounding concerns any factor that creates a positive correlation between having a connection and the student's matching outcome—but only compared to the student's peers in the same class and only to the extent that the factor increases the probability that the student matches with a given university even if their advisor did not have a connection there. One such example is a case where there is ethnicity-based homophily for collaboration connections—both for the graduate-advisor and for the advisor-coauthor relationship—together with an ethnicity-specific preference for locations: Students would be more likely to match with universities where their advisor has a co-author, but this would occur even if the advisor did not have a connection.

Assessing changes over time The stark rise in the number of connections to universities highlights the possibility that the effect of a given connection changes over time. Mechanically, if the number of connections increases over time, the probability of a match through a given connection decreases at some point because graduates can only go to one university after graduation. However, the overall probability of a match within the set of network connections can increase over time. To test for changes over time in γ and to not confound differences over time with changes in the composition of research fields, we estimate the following regression model

$$E_{i,j} = \alpha_{c(i),j} + \beta X_{i,j} + \underbrace{\sum_{F} \sum_{n=0}^{2} \gamma_F^n A_{i,j} N_i^n \mathbf{1}\{\text{Field}_i = F\}}_{\text{Heterogeneity: Field + N connection}} + \underbrace{\delta A_{i,j} \times (t(i) - 1990)}_{\text{linear time trend}} + \varepsilon_{i,j}. \quad (2)$$

We test whether $\delta = 0$ while accounting for heterogeneity in network effects across fields and the number of connections. Specifically, we estimate a field-specific quadratic function in the number of connections N_i. The number of connections of the student to the potential hiring university is measured separately for PhD graduates' own co-author connections and their advisors' co-author connections.

Post-PhD outcomes of network hires We compare the careers of PhD graduates with a first affiliation where a network connection is present to those without a network connection. To compare post-PhD outcomes $y_{i,j(i)}$ we use the following regression framework

$$y_{i,j(i)} = \exp\left(\underbrace{\alpha_{c(i)}}_{\text{class effect}} + \underbrace{\alpha_{j(i)}}_{\text{destination effect}} + \underbrace{\beta X_i}_{\text{individual controls}} + \underbrace{\gamma A_{i,j}}_{\text{network effect}} + u_i \right), \tag{3}$$

where $A_{i,j}$ indicates whether a graduate's first affiliation post-PhD has been at a university with a pre-graduation network connection.

Our main interest is in the parameter γ, which measures the difference in the post-PhD outcomes between graduates placed through the network and graduates not placed through the network. By varying which fixed effects, we include when estimating equation (3), we study the role of the network from the point of view of graduates and hiring universities. First, we compare graduates in the same class with a class fixed effect $\alpha_{c(i)}$. Second, we compare new hires at the same university with a first affiliation fixed effect $\alpha_{j(i)}$—this specification relates to studies using productivity outcomes at a few firms (Burks et al., 2015) but for all employers in the market. Third, we combine the preceding specifications and include both types of fixed effects.

We also assess the role of the students' observable characteristics X_i. These include the position of the graduate's pre-graduation and the advisor's publications in their respective citation distributions within a 5-year interval and research fields.[3]

The universities also observe similar information at the time of hiring; our regressions with and without these controls will inform whether networks provide any additional

[3]See the appendix for details.

information beyond these variables. This closely relates to the results in existing work with skill measures (Burks et al., 2015; Pallais and Glassberg Sands, 2016). While similar, there are two key differences to those earlier results. First, we observe an occupation-specific productivity measure. Second, we observe this measure also for the advisor who plays an important role in the referral process—instead of only the graduate's pre-hiring productivity signals.

Assessing changes over time We investigate whether the selection on ex-post outcomes between connected and not connected graduates changed during our sampling period by expanding the framework in equation (3):

$$y_{i,j(i)} = \exp\left(\alpha_{c(i)} + \alpha_{j(i)} + \beta X_i + \sum_F \gamma_F A_{i,j} \mathbf{1}\{F = \text{Field}_i\} + \underbrace{\delta A_{i,j} \times (t(i) - 1990)}_{\text{linear time trend}} + u_i\right), \tag{4}$$

We are interested in the parameter δ, which estimates a linear trend in the outcome gap between connected and non-connected graduates. The specification also accounts for possible confounding arising from differences across research fields: First, it accounts for differences in the gap between connected and non-connected graduates across fields (captured by field-specific γ_F); second, X_i includes fixed effects at the level of the graduate's research field $\times$ graduating year, accounting for field-specific time trends in the post-PhD research outcomes during our sampling period.

Results

We now present the main results of our analysis. First, we estimate the effect of the advisor's co-author connections on the student's first academic placement. Second, we compare the post-PhD career outcomes between graduates who were hired with a connection present and graduates who were hired without a connection present. Throughout the analysis, we discuss the results for the advisor's connections but also report results for PhD graduates' own connections.

The impact of network connections on the first affiliation of PhD Graduates

Table 2 shows that the co-authorship connections of the PhD advisor impact the matching of PhD graduates to universities. We estimate a linear probability model, where each observation is a pair of a PhD graduate and a potential hiring university. The first column shows the raw average probability that a graduate matches with a given university. It is 0.475pp for a graduate-university pair without any network connection and $0.475+1.15 = 1.625$pp for a graduate-university pair with an advisor connection.

Since graduates have, on average, more than ten connections to other universities through the advisor, this suggests a substantial role for the advisor's network in matching PhD graduates to their first post-PhD affiliation. However, these differences may be driven by factors other than the connection itself, and we now tighten the regression models.

In column 2, we add fixed effects for the class $\times$ potential hiring university pair. Comparing within these pairs, graduates whose advisors have a connection to a given university are 0.599pp more likely to be hired by that university than peers whose advisors do not have a connection to that university. This specification controls for any systematic matching determinants that vary at the level of the PhD class $\times$ potential hiring university pair, for example, geographic distance or research specialization within major fields at the university level (e.g. specialization into plant biology).

Controls for the research productivity of the PhD student before graduating and that of the advisor—added in columns 3 and 4—leave the estimated coefficients almost unchanged and confirm that the effects are neither spuriously driven by the prominence of the advisor in the field nor the student's pre-graduation productivity. In column 5, we include a fixed effect for each PhD graduate, thus comparing options only within a

student's choice set. Again, the estimates are not substantially altered.

Table 2: Network Effects: Matching of PhD Graduates to their first post-PhD university

Panel A: Average effect of network connections on placement						
Dependent Variable:			Match formed			
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
Constant	0.475					
	(0.021)					
Advisor connection	1.15	0.599	0.598	0.604	0.649	0.603
	(0.102)	(0.037)	(0.036)	(0.036)	(0.038)	(0.035)
PhD's connection	2.85	3.31	3.30	3.30	3.37	3.31
	(0.174)	(0.152)	(0.149)	(0.149)	(0.155)	(0.153)
Fixed-effects						
PhD Class×Potential Hiring University ID		Yes	Yes	Yes	Yes	Yes
Pre-Graduation Productivity×Field×5 Year Window			Yes	Yes		
Advisor Citation Decile×Field×5 Year Window				Yes		
Student Id					Yes	Yes
Additional controls with varying slopes						
Max similarity to faculty members×Field						Yes
Avg. similarity to faculty members×Field						Yes
Observations	5,396,046	5,396,046	5,396,046	5,396,046	5,396,046	5,396,046

Panel B: Assessing changes over time in effect of network connections			
Dependent Variable:		Match formed	
Model:	(1)	(2)	(3)
Variables			
Advisor connection	0.885		
	(0.069)		
Advisor connection $\times(t-1990)$	-0.017	-0.018	-0.006
	(0.003)	(0.003)	(0.003)
Fixed-effects			
PhD Class×Potential Hiring University ID	Yes	Yes	Yes
Student Id	Yes	Yes	Yes
PhD's connection	Yes		
PhD's connection × Field		Yes	Yes
Advisor connection × Field		Yes	Yes
Additional controls (with varying slopes)			
Max similarity to faculty members × Field	Yes	Yes	Yes
Avg. similarity to faculty members × Field	Yes	Yes	Yes
PhD's connection $\times(t-1990)$	Yes	Yes	Yes
Advisor connection$\times N\times$ Field			Yes
PhD's connection$\times N\times$ Field			Yes
Advisor connection$\times N^2\times$ Field			Yes
PhD's connection$\times N^2\times$ Field			Yes
Observations	5,396,046	5,396,046	5,396,046

Notes: Unit of observation is a pair of PhD graduate and a potential hiring university. The sample is restricted to pairs of a PhD class (PhD university×Field×5 year window) and potential hiring university with variation in the connection status of PhD graduates. See equation (1) for the regression specification for Panel A and equation (2) for Panel B. Clustered (PhD university×Field×5 year window + potential hiring university) standard-errors in parentheses

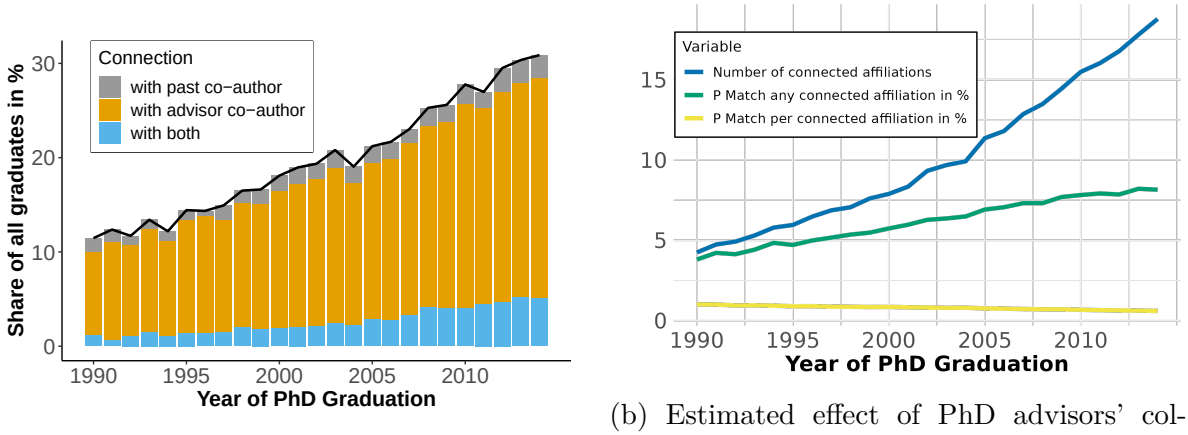
So far, we have accounted for potential confounders that vary at the level of the interaction of PhD class and potential hiring university. However, an individual PhD graduate's specialization into subfields is related to both the probability that their advisor

is connected to a given university and whether the student would be hired in the absence of the connection. To address this, we show in column 6 that the PhD graduate's specialization into subfields does not drive the measured effect of network connections on matching. We add controls for the cosine similarity of topics of the PhD graduate dissertation to the potential hiring university research output. We separately control for the maximum similarity to a faculty member and the average similarity to all research output at the potential hiring university within the same field. While these measures predict which university a graduate matches with, they are not spuriously driving the measured effect of network connections on the matching of PhD graduates and universities.

To sum up our results so far, we find that pre-existing network connections, both through the advisors' co-author network and through the PhD graduate's own co-authors, have a substantial effect on the matching of PhD graduates to their first university after graduating. This holds even when comparing employment options only within a given PhD graduate's choice set and controlling for the overlap in research topics of the PhD graduate's dissertation and the potential hiring university. A connection to a particular university through the advisor increases the probability of matching with that university by about 0.6pp, more than doubling the probability of matching. The effect of the advisor's collaboration network is substantial overall, as a PhD graduate is connected to over ten universities through her advisor on average.

In Panel B of table 2, we show the estimated time trend in the advisor's co-author network's effect on matching PhD graduates to their first university. The average connection has a lower effect on the matching probability in more recent years, as shown in column (1). In column (2), we add controls for differences in the effect of connections by field, and in column (3), we further allow the effect of connections to vary with the number of connections. The coefficient on the time trend in column (3) is modest (-0.006, s.e. 0.003).

In Figure 1b, we show that the PhD advisors' collaboration network has become more important for placing PhD graduates between 1990 and 2014. This is driven by the rise in the number of connections that PhD graduates have through their advisors, as



(a) Share of PhD graduates placed within collaboration network

(b) Estimated effect of PhD advisors' collaboration network on placement of PhD graduates

Figure 1: Share of hires in Collaboration Network and its estimated effect on placement

Notes: In Panel A the share of graduates with a pre-existing collaboration network connection to their first affiliation are shown. The sample is the same as in column (2) of table 1. In Panel B we show the average number of connections in blue. The yellow line shows the average of the estimated effect of one advisor connection on the matching probability. Specifically, it is $\hat{\gamma}_i$ based on the estimation of equation (2) done separately by field. For each PhD graduate, $\hat{\gamma}_i = \hat{\gamma}^0_{\text{Field}_i} + \hat{\gamma}^1_{\text{Field}_i} N_i + \hat{\gamma}^2_{\text{Field}_i} N_i^2 + \delta_{\text{Field}_i}(t-1990)$, where N_i is the number of connections the PhD advisor has to potential hiring universities. The green line shows the cumulative effect of all connections on the matching probability, the average of $\hat{\gamma}_i N_i$.

shown by the blue line, which rose from approximately 5 in the early 90s to over 15 by 2010. The effect of any single connection on the matching probability has declined from approximately 1pp to 0.6pp, as shown by the yellow line. Therefore, the cumulative effect of all connections, indicated by the green line, has risen from 4pp in the early 90s to 8pp in the 2010s.

Post-PhD outcomes of network hires

We now study whether graduates placed through their advisor's pre-existing co-author network have different career outcomes after graduation compared to graduates not placed through the advisor's network. We focus on outcomes measuring research productivity, such as citations and the number of papers published in the first 7 years after graduation. We complement these results with additional outcomes on (i) collaboration patterns, measured by the number of new co-authors at the first post-PhD affiliation, (ii) stability of the match, measured by an indicator for whether a graduate is still affiliated with the same university 6 years after their PhD, (iii) whether graduates produce any output more

than 6 years after the PhD and (iv) the productivity of the hiring university measured by citations.

The results are shown in Table 3, separately for the different comparisons: Panel A shows the results from the point of view of the graduating class, Panel B from the point of view of the hiring university, and Panel C includes fixed effects for both. All results refer to poisson regressions so that the coefficients measure differences in log points. We estimate the regressions with the *fixest* package in `R` (Bergé, 2018).

Comparing within the graduating class Starting with panel A, we find that graduates placed through connections are more productive than their classmates not placed through connections. Column 1 shows that this gap is 26.2 log points for the number of citations.[4] This gap can reflect pre-determined ability, access to productivity-enhancing employers, or a direct effect of the connection as such.

Columns 2 and 3 show that pre-determined and publicly observable predictors of productivity can only partly explain this gap. Controlling for the expected citations of the graduate's work before graduation lowers the point estimate to 0.186; further controlling for the advisor's citations before graduation lowers the estimate to 0.139 (standard error 0.021). Thus, graduates from the same PhD with a connection to their first post-PhD university outperform, above and beyond, what is predictable based on their own and their advisor's pre-graduation research output. In the remaining columns of the table, we include these additional predictors as controls in the regressions.

Column 4 shows that a higher number of papers partly drives higher citations. The point estimate is 0.061 (standard error 0.01), indicating that less than half the higher number of citations post-PhD is due to more papers. Thus, graduates placed through connections are more productive in the quantity and quality of research.

Column 5 shows that graduates placed through connections collaborate more with their new colleagues. Despite increased collaboration, connected hires are more than 13% less likely to be affiliated with their first university 6 years after the PhD (column 6). This result is not driven by dropping out from research: Column 7 shows that the

[4]The point estimate translates into $exp(0.262) - 1 \approx 30\%$ higher citations.

probability of producing any output 6 years after the PhD or later is the same.

Finally, in column 8 we show that graduates placed through connections are matched to substantially more productive universities, as measured by the citations of the university in the same major field.

To summarize, from the perspective of the PhD class, graduates placed through the network have substantially different careers compared to their peers. They outperform post-PhD in both quantity and quality of research, are placed in more productive universities, collaborate more with their new colleagues, and are more likely to switch affiliations 6 years after the PhD. This holds even after controlling for pre-determined, publicly observable predictors of productivity. In the appendix, we also report similar results when comparing connected and not connected graduates with the same advisor.

Panel A: Comparison of post-PhD outcomes with Class Fixed Effect

Dependent Variables:	N Cites PhD graduate			N papers	Co-authors First Affil	Same Affil PhD+6yrs	Any output PhD+6yrs	N Cites of First Affil
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variables								
Advisor connection	0.262	0.186	0.139	0.061	0.199	-0.143	-0.005	0.599
	(0.023)	(0.019)	(0.021)	(0.010)	(0.021)	(0.013)	(0.004)	(0.101)
PhD's connection	0.215	0.034	0.048	0.069	-0.086	-0.588	-0.024	-0.028
	(0.029)	(0.027)	(0.027)	(0.012)	(0.022)	(0.029)	(0.006)	(0.025)
Fixed-effects								
PhD Class	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Subfield (MAG lvl 1)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pre-Graduation Productivity×Field		Yes	Yes	Yes	Yes	Yes	Yes	Yes
Advisor Citation Decile×Field			Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics								
Pseudo R^2	0.44	0.51	0.52	0.30	0.36	0.04	0.009	0.76
Observations	73,775	73,775	73,775	73,885	69,566	71,704	73,427	73,705

Panel B: Comparison of post-PhD outcomes with Destination Fixed Effect

Dependent Variables:	N Cites PhD graduate			N papers	Co-authors First Affil	Same Affil PhD+6yrs	Any output PhD+6yrs
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Variables							
Advisor connection	0.129	0.058	-0.013	0.006	0.026	-0.065	-0.009
	(0.020)	(0.018)	(0.018)	(0.009)	(0.015)	(0.012)	(0.005)
PhD's connection	0.191	0.026	0.048	0.059	-0.077	-0.559	-0.022
	(0.026)	(0.024)	(0.024)	(0.011)	(0.018)	(0.029)	(0.006)
Fixed-effects							
Field×5 Year Window	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Hiring University Id×Field	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Subfield (MAG lvl 1)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pre-Graduation Productivity×Field		Yes	Yes	Yes	Yes	Yes	Yes
Advisor Citation Decile×Field			Yes	Yes	Yes	Yes	Yes
Fit statistics							
Pseudo R^2	0.41	0.48	0.49	0.30	0.35	0.04	0.008
Observations	73,672	73,672	73,672	73,885	66,819	72,008	73,273

Panel C: Comparison of post-PhD outcomes with Class and Destination Fixed Effect

Dependent Variables:	N Cites PhD graduate			N papers	Co-authors First Affil	Same Affil PhD+6yrs	Any output PhD+6yrs
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Variables							
Advisor connection	0.091	0.041	-0.008	0.018	0.035	-0.059	-0.006
	(0.021)	(0.020)	(0.019)	(0.010)	(0.018)	(0.013)	(0.005)
PhD's connection	0.174	0.012	0.026	0.050	-0.090	-0.571	-0.022
	(0.028)	(0.027)	(0.027)	(0.013)	(0.022)	(0.030)	(0.006)
Fixed-effects							
PhD Class	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Hiring University Id×Field	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Subfield (MAG lvl 1)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pre-Graduation Productivity×Field		Yes	Yes	Yes	Yes	Yes	Yes
Advisor Citation Decile×Field			Yes	Yes	Yes	Yes	Yes
Fit statistics							
Pseudo R^2	0.54	0.60	0.61	0.37	0.42	0.07	0.014
Observations	73,596	73,596	73,596	73,885	64,634	70,109	72,893

Notes: Unit of observation is a PhD graduate. See equation (3) for the poisson regression specification. Observations with zero outcomes that are perfectly predicted by fixed effects are dropped. Clustered (PhD university-field-5 year window + Hiring university) standard-errors in parentheses. N Cites PhD graduate measures citations received on articles published in the first 6 years post PhD graduation. N papers is the number of articles published in the same period. Co-authors First affil measures the number of new co-authors at the first post-PhD affiliation. Same Affil PhD+6yrs and Any Output PhD+6yrs indicate whether the PhD graduate, 6 years after the PhD graduation or later, is still affiliated with their first post-PhD affiliation and whether they publish any papers at that point.

Table 3: Post-PhD outcomes of connected vs. not connected hires

Comparing within the hiring university In panel B, we show the results from the point of view of the hiring university. Column 1 shows that connected hires are about 13% more productive than other hires at the same university. Thus, connected hires are positively selected in terms of productivity.

In columns 2 and 3, we progressively add the pre-determined productivity predictors as in panel A, showing that publicly observable pre-determined factors can explain the gap. Including both the advisor's and the student's pre-graduation productivity (in column 3) makes connected and non-connected hires indistinguishable—the point estimate for γ drops to -0.013 (standard error 0.018). This means that connected hires are positively selected in terms of publicly observable productivity predictors at the time of PhD graduation but not in terms of private information revealed by the connection. Therefore, networks do not reveal private information about productivity on either side of the market, but they allow departments to hire PhD graduates who are expected to be more productive.

The remaining columns in panel B keep the same controls as column 3. Compared to other hires at the same university, they show that connected hires publish a similar number of papers and collaborate with a similar number of new co-authors but are more likely to leave the university despite continuing research at a similar rate.

The results about the role of publicly observable predictors of productivity in panel B are noteworthy for three reasons. First, they contrast with the findings from panel A from the graduating class' point of view: Connected hires from a similar background substantially outperform their non-connected peers, but once the place of work is taken into account, this gap drops to 0. This implies that either the place of work has a direct effect on the productivity of graduates or that the hiring mechanism of these more productive places can uncover information about graduates that is unobserved to the analyst, also in the absence of a network connection. Second, the gap in column 1 resembles the findings in Burks et al. (2015) for trucking and high-tech. But, while they found that connected hires were similar along measures of general human capital (Burks et al., 2015, Table IV), we find that the gap can be explained by a more precise

measure of pre-determined productivity predictors specific to the job at hand. Third, the results are not in line with theories that highlight the use of networks to reveal information about match quality: Such theories imply that connected hires should be more productive than non-connected hires even after controlling for any publicly observable predictors of productivity. We find no evidence for such effects. Additionally, connected hires are likelier to move to a different university 6 years after earning their PhD. To the extent that separations depend on the quality of the match, these results also suggest that the network reveals information about match quality neither in terms of productive nor in terms of non-productive factors.

Comparing within class and within hiring university Panel C in Table 3 shows results where we control for fixed effects for the hiring university and the PhD class. The results are similar to those in panel B: Connected hires outperform other hires at the same university, but this performance gap can be predicted by public information at the time of hiring. Point estimates are very similar when adding the fixed effect for the PhD class in addition to the hiring university. This implies that conditional on the selection done by the hiring university, there is limited additional information in knowing which PhD class a student is from when comparing connected to not connected hires. The evidence implies that the net effect of network hiring on productivity, either through revealing private information or through direct productivity effects, is approximately zero. The 95% confidence interval covers values between -4.5% and $+3\%$. While network hiring allocates graduates to more productive universities on average, the evidence suggests their productivity would have been similar without the network connection if working at the same place.

Assessing Changes over time In the Online Appendix, we report the results from estimating equation (4), with the fixed effects for both class and field. We are unable to detect substantial changes in the difference in ex-post hiring outcomes between graduates hired through the network and graduates not hired through the network. Thus, despite the increasing importance of the collaboration network for PhD graduates' hiring, the

selection into being hired through the network has remained stable during our sampling period.

Conclusion

We analyze the role of PhD advisors' collaboration networks for matching PhD graduates to universities. To do so, we build a novel database containing information about PhD graduates that allows us to track them and their advisors throughout their careers.

We find that pre-existing collaboration connections of advisors strongly predict which affiliation a PhD graduate matches to—having a connection more than doubles the probability of matching with a given university. Our estimates account for unobserved heterogeneity at the student level, at the sending $\times$ hiring university pair level, and for the similarity in research topics between the student's dissertation and her potential new colleagues at the destination university. Overall, the importance of advisors collaboration networks for matching PhD graduates to universities doubled between 1990 and 2014 due to the increasing size of the network.

We then show whether there is a gap between connected and non-connected graduates on a range of post-hiring performance measures. When comparing graduates within the same class, connected graduates are more productive. Even after controlling for the research output of the graduate and her advisor during the PhD connected graduates receive over 14% more citations. When comparing hires at the same university, connected hires are more productive. However, this productivity gap can be predicted by public information at the time of hiring—namely, by the research output of the graduate and her advisor during the PhD.

These results indicate that on average, an advisor's network is used to the benefit of both the PhD graduates and the hiring universities. From the point of view of the hiring university, the network helps in hiring more productive graduates, but no private information is revealed beyond what is publicly observable at the time of hiring. From the point of view of the graduating class, graduates placed through the network are more

productive and work at more productive universities than graduates not placed through the network; this holds even when comparing graduates with the same advisor. This highlights a crucial role for advisors selectively using their network to allocate graduates who are more productive to more productive universities.

Lastly, we find no evidence for increased productivity of connected hires after accounting for observable information and the selection done by the hiring university. This suggests that network hiring has no aggregate productivity effects.

References

Anderson, Katharine A, and Seth Richards-Shubik. 2021. "Collaborative production in science: An empirical analysis of coauthorships in economics." *The Review of Economics and Statistics*, 1–45.

Bergé, Laurent. 2018. "Efficient estimation of maximum likelihood models with multiple fixed-effects: the R package FENmlm." *CREA Discussion Papers*, , (13).

Brown, David G. 1965. *Academic labor markets.*

Brown, Meta, Elizabeth Setren, and Giorgio Topa. 2016. "Do informal referrals lead to better matches? Evidence from a firm's employee referral system." *Journal of Labor Economics*, 34(1): 161–209.

Burks, Stephen V, Bo Cowgill, Mitchell Hoffman, and Michael Housman. 2015. "The value of hiring through employee referrals." *The Quarterly Journal of Economics*, 130(2): 805–839.

Clauset, Aaron, Samuel Arbesman, and Daniel B Larremore. 2015. "Systematic inequality and hierarchy in faculty hiring networks." *Science advances*, 1(1): e1400005.

Dustmann, Christian, Albrecht Glitz, Uta Schönberg, and Herbert Brücker. 2016. "Referral-based job search networks." *The Review of Economic Studies*, 83(2): 514–546.

Eliason, Marcus, Lena Hensvik, Francis Kramarz, and Oskar Nordström Skans. 2022. "Social connections and the sorting of workers to firms." *Journal of Econometrics.*

Fortunato, Santo, Carl T Bergstrom, Katy Börner, James A Evans, Dirk Helbing, Staša Milojević, Alexander M Petersen, Filippo Radicchi, Roberta Sinatra, Brian Uzzi, et al. 2018. "Science of science." *Science*, 359(6379): eaao0185.

Freeman, Richard B, Ina Ganguli, and Raviv Murciano-Goroff. 2014. "Why and wherefore of increased scientific collaboration." In *The changing frontier: Rethinking science and innovation policy.* 17–48. University of Chicago Press.

Galenianos, Manolis. 2013. "Learning about match quality and the use of referrals." *Review of Economic Dynamics*, 16(4): 668–690.

Gregg, Forest, and Derek Eder. 2022. "Dedupe, version 2.0.11. URL: https://github.com/dedupeio/dedupe."

Hadlock, Charles J, and Joshua R Pierce. 2021. "Hiring your friends: Evidence from the market for financial economists." *ILR Review*, 74(4): 977–1007.

Hensvik, Lena, and Oskar Nordström Skans. 2016. "Social networks, employee selection, and labor market outcomes." *Journal of Labor Economics*, 34(4): 825–867.

Huang, Junming, Alexander J Gates, Roberta Sinatra, and Albert-László Barabási. 2020. "Historical comparison of gender inequality in scientific careers across countries and disciplines." *Proceedings of the National Academy of Sciences*, 117(9): 4609–4616.

Jones, Benjamin F. 2009. "The burden of knowledge and the "death of the renaissance man": Is innovation getting harder?" *The Review of Economic Studies*, 76(1): 283–317.

Kramarz, Francis, and David Thesmar. 2013. "Social networks in the boardroom." *Journal of the European Economic Association*, 11(4): 780–807.

Kramarz, Francis, and Oskar Nordström Skans. 2014. "When strong ties are strong: Networks and youth labour market entry." *Review of Economic Studies*, 81(3): 1164–1200.

Long, J Scott, and Robert McGinnis. 1985. "The effects of the mentor on the academic career." *Scientometrics*, 7: 255–280.

Montgomery, James D. 1991. "Social networks and labor-market outcomes: Toward an economic analysis." *The American economic review*, 81(5): 1408–1418.

Pallais, Amanda, and Emily Glassberg Sands. 2016. "Why the referential treatment? Evidence from field experiments on referrals." *Journal of Political Economy*, 124(6): 1793–1828.

Proquest. 2021. "Proquest Dissertation & Theses Global. `https://about.proquest.com/en/products-services/pqdtglobal/` (accessed 2021)."

Rajkumar, Karthik, Guillaume Saint-Jacques, Iavor Bojinov, Erik Brynjolfsson, and Sinan Aral. 2022. "A causal test of the strength of weak ties." *Science*, 377(6612): 1304–1310.

Rees, Albert. 1966. "Information networks in labor markets." *The American Economic Review*, 56(1/2): 559–566.

Rose, Michael E, and Suraj Shekhar. 2023. "Adviser Connectedness and Placement Outcomes in the Economics Job Market." *Labour Economics*, 102397.

Simon, Curtis J, and John T Warner. 1992. "Matchmaker, matchmaker: The effect of old boy networks on job match quality, earnings, and tenure." *Journal of labor economics*, 10(3): 306–330.

Sinha, Arnab, Zhihong Shen, Yang Song, Hao Ma, Darrin Eide, Bo-June (Paul) Hsu, and Kuansan Wang. 2015. "An Overview of Microsoft Academic Service (MAS) and Applications. `https://zenodo.org/record/2628216` (accessed 2021)." 243–246.

Terviö, Marko. 2011. "Divisions within academia: Evidence from faculty hiring and placement." *Review of Economics and Statistics*, 93(3): 1053–1062.

Topa, Giorgio. 2011. "Chapter 22 - Labor Markets and Referrals." In . Vol. 1 of *Handbook of Social Economics*, , ed. Jess Benhabib, Alberto Bisin and Matthew O. Jackson, 1193–1221. North-Holland.

Wang, Kuansan, Zhihong Shen, Chiyuan Huang, Chieh-Han Wu, Darrin Eide, Yuxiao Dong, Junjie Qian, Anshul Kanakia, Alvin Chen, and Richard Rogahn. 2019. "A review of microsoft academic services for science of science studies." *Frontiers in Big Data*, 2: 45.

Wang, Kuansan, Zhihong Shen, Chiyuan Huang, Chieh-Han Wu, Yuxiao Dong, and Anshul Kanakia. 2020. "Microsoft academic graph: When experts are not enough." *Quantitative Science Studies*, 1(1): 396–413.

Wuchty, Stefan, Benjamin F Jones, and Brian Uzzi. 2007. "The increasing dominance of teams in production of knowledge." *Science*, 316(5827): 1036–1039.