# Networks in the market for researchers*

Flavio Hafner†
Christoph Hedtrich‡

First version: October 4, 2022
This draft: September 14, 2023

**Abstract**

We study the role of networks in the labor market for young scientists in the United States. Nearly one in five PhD graduates that publish after PhD graduation do so at a university where their advisor has a former co-author; graduates that have such a connection are more than twice as likely to match with the university, even within fine-grained peer groups. We document a citation premium of 12 to 36% for graduates placed through the advisor's network. However, we find no evidence for private information about productivity being revealed through the network.

*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 rely heavily on finding a match through their social connections (Topa, 2011). One possible reason 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 can provide valuable information that lowers the search cost. But whether this leads to better matches is ambiguous. In the optimistic view, referrers promote social contacts that are a good fit 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 social contacts, leading to nepotism and inefficient matches.

This tension is particularly salient in the market for scientists because agency problems stemming from repeated co-authorship interactions may be severe (Caplow and McGee, 1958; Hargens and Hagstrom, 1967). The labor market for scientists also provides a unique setting to study the role of networks for matching and productivity—with precise data on workers' pre- and post-hiring productivity, on thesis advisors as potential referrers and their networks, and on match-specific factors such as the research topics of candidates and employers.

We harness this setting for three purposes: to show the role of social connections in hiring and productivity outcomes for job candidates from similar backgrounds; to assess whether social connections reveal private information about either the candidate or match quality at the time of hiring, or whether they serve mainly to attract job candidates; and in turn, show whether the use of social connections in hiring is associated with higher productivity overall.

Existing studies addressing these questions face important data constraints. They 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 often can not directly measure network connections for a large sample (Dustmann et al., 2016).

The main contribution of our paper is to overcome these data constraints by focusing on young scientists. This allows us to contribute in two ways. First, we observe a close approximation to each graduate's choice set of universities where they could find their first job. When studying hiring flows, we can therefore control for important potential confounders: we account not only for any bilateral determinants of hiring flows between graduating and hiring universities but also for the fit in terms of research topics between graduate and potential employer. Second, we provide new results on the productivity of network hires. We are able to compare graduates with very similar backgrounds and networks. Thus, we can account for selection into having network connections, which otherwise may confound the comparison of network and non-network hires. Further, we study how publicly observable pre-hiring productivity predictors explain differential research output of network and non-network hires to distinguish the importance of different theories of network hiring.

We start our analysis by showing that the advisor's collaboration network has a big impact on where students find their first job. Our empirical design follows Eliason et al. (2022) and compares graduates with and without a network connection from the same background—in our case the PhD university, field, and year of graduating—and at the same potential hiring university while controlling for additional individual determinants of both networks and potential match quality.

We estimate that having a connection to a university through the advisor's network more than doubles the probability that the student is matched to that university. The results are precisely estimated and robust across specifications where we successively compare employment options across students from the same class; compare employment options within the same student; and account for how similar the student's research topics are to those of the potential hiring university.

These 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: Existing work focuses on hiring networks between universities (Terviö, 2011; Clauset, Arbesman and Larremore, 2015; Anderson and Richards-Shubik, 2021); in contrast, our unit of analysis are the networks of individual researchers, and we document an underlying channel for imbalanced hiring flows between universities in the aggregate. The findings are also in accord with the notion that the PhD advisors and their connections play a crucial role in matching students to their first job (Long and McGinnis, 1985).

Our estimates of the effect of the advisor's network on matching have a causal interpretation under the assumption that there are no factors that simultaneously determine where the advisor has a network connection and where the graduate finds their first job, relative to the peers from the same graduating class. Importantly, these factors need to be unobserved to the analyst but observed by the student while on the job market *even in the absence of an actual network connection.* This assumption is not testable, but we note that 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

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

ex-post more output than non-connected hires. This is in line with existing work with output data—Pallais and Glassberg Sands (2016) and the high-skill setting in Burks et al. (2015). Our results show that, relative to non-connected hires, connected hires generate more citations, write more papers, and collaborate more with their new colleagues.

By digging deeper into the data on graduates' research output, 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 pre-match productivity and the identity of the referrer.

When we compare graduates within the same class, we find that those who had a pre-existing connection to their new university produce 36% 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 also write more papers, collaborate more with their new colleagues, and are placed at more prestigious universities. We find very similar results in specifications with advisor fixed effects, indicating that advisors 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 who had a pre-existing connection produce 17% 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 small 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 hiring 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, with a few exceptions, the results are not in line 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 workers that are more productive 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. This implies that the network does not reveal private information about students' productivity but instead suggests that networks lower hiring costs, 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, 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 and paper titles as well as starting their publication career around to 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 in more detail the data processing and the quality of the data. First, both data sources are of high quality. ProQuest is used by the US Library of Congress; 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.

Our sample consists of PhD graduates from US universities that received their degree between 1990 and 2015 and whom we linked to an author in MAG. We restrict the sample

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

to graduates whom 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 for whose advisor we have linked to MAG.

For each graduate, we construct their connections to universities through their own past collaboration network and through the network of their advisor. We also compare the similarity in research concepts between the student and their potential new colleagues. To do so, we compute the cosine similarity between the dissertation concepts and the concepts of papers published by faculty at the potential destination university. We use the language model provided by (Wang et al., 2020) to predict topics based on abstract text for both dissertations and faculty research. Thus we use a consistent definition of concepts based on the same corpus.

To do so, we extract concepts from paper abstracts. For papers in MAG, the concepts are available in the data. For dissertations, we apply the same language model provided with MAG (Wang et al., 2020).[3]

Table 1: Summary statistics

| Variable | Baseline | Advisor in MAG | Connected first affiliation | Not Connected first affiliation |
|---|---|---|---|---|
| Degree Year | 2004 | 2005 | 2006 | 2004 |
| Year First pub post PhD | 2.9 | 2.9 | 2.5 | 2.9 |
| N Cites 7y post PhD | 352.1 | 347.4 | 555.2 | 288.8 |
| N Papers 7y post PhD | 11.3 | 11.3 | 14.3 | 10.5 |
| Link Score Student | 0.97 | 0.98 | 0.97 | 0.98 |
| Link Score Advisor | | 0.97 | 0.97 | 0.97 |
| Connected Advisor | | 0.20 | 0.92 | 0 |
| Connected Own Co-author | | 0.05 | 0.22 | 0 |
| N Affiliation Connections | | | | |
| through advisor | | 12.7 | 23.0 | 9.8 |
| through co-author | | 2.3 | 4.3 | 1.7 |
| N PhD Graduates | 92,490 | 77,192 | 16,979 | 60,213 |
| N PhD Graduates × | | | | |
| potential hiring institutions | 34,281,730 | 29,047,573 | 6,269,981 | 22,777,592 |

*Notes:* The baseline sample is PhD graduates in the Proquest Dissertations&Theses database from US universities included in the Carnegie Classification between 1990 and 2015 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 main analysis sample. Columns 3 and 4 split this sample into those who have a network connection to their first affiliation and those who do not.

The baseline sample of graduates linked to MAG consists of 92,490 graduates, for 77,192 of whom we have advisor information and were able to link the advisor to MAG. This sample is our main 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.

[3] Details are discussed in Wang et al. (2020, section 2.5)

In the main 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 the matching of PhD graduates to their first post-PhD affiliation. It also resonates with early evidence in Brown (1965), who found that about 12% of newly hired faculty found a job through their advisor.

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. We will return to this pattern in the empirical section 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 follow Eliason et al. (2022) and use the following regression framework

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

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 $\gamma$. 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. As an 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 the matching of 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 students from the same class that are specialized 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 both 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 kinds of confounders are not captured by our controls and which could invalidate a causal interpretation of the network parameter $\gamma$. 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 example is that students' taste for particular universities correlates with factors that lead their advisor to choose particular co-authors. When considering ethnicity, for example, it may be that Chinese researchers work particularly well with other Chinese researchers due to the common native language, so Chinese advisors have more Chinese co-authors and more Chinese students. At the same time, Chinese researchers want to live on the West Coast because their trips to visit relatives are shorter. Thus, there are more co-authors of the advisor on the West Coast, and students are more likely to match with the same universities—but this would occur even if the advisor did not have a connection.

**Heterogenous network effects** To incorporate heterogenous network effects into the above framework, define $\gamma$ and $A_{i,j}$ as vectors, where each element corresponds to a different network type. For instance, we include the PhD graduate's own co-author connections as a control alongside the advisor's co-author connections. Then, $\gamma A_{i,j} = \sum_n \gamma^n A_{i,j}^n$ is the sum of the network effects of all network types $n$. See Eliason et al. (2022) for a more detailed description of the estimation of heterogenous network effects.

**Post-PhD outcomes of network hires** We compare the careers of PhD graduates who have 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

$$y_{i,j(i)} = \exp\left( \underbrace{\alpha_{c(i)}}_{\text{class effect}} + \underbrace{\alpha_{j(i)}}_{\text{destination effect}} + \underbrace{\beta X_i}_{\text{individual controls}} + \underbrace{\gamma A_{i,j}}_{\text{network effect}} + u_i \right), \quad (2)$$

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

Our main interest is in the parameter $\gamma$, 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 (2), 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 $X_i$. The first measure is the expected citations of the student's pre-graduation publications. We predict the number of citations with the average citations of papers published in the same year and venue (conference or journal).[4] Then, we compute the student's position in the distribution, relative to other students in the same research field and within a five-year interval. The second measure is the advisor's position in the citation distribution in the last 10 years, within the respective research field, and a five-year interval before graduation.

These variables are observed by the universities at the time of hiring; our regres-

[4] We consider this measure more precise than the acutal citations received by the time of graduation.

sions with and without these controls will inform about whether networks provide any additional information beyond these easily observed 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.

## 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 to 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 0.477pp for a graduate-university pair without any network connection and $0.477 + 1.16 = 1.647$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 the matching of PhD graduates to their first post-PhD affiliation. However, these differences may be driven by other factors 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 advisor has a connection to a given university are 0.598pp more likely to be hired by that university, compared to peers whose advisor does 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.

So far we have accounted for potential confounders that vary at the level of a pair 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.

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

| Dependent Variable: | Match formed | | | | | |
|---|---|---|---|---|---|---|
| Model: | (1) | (2) | (3) | (4) | (5) | (6) |
| *Variables* | | | | | | |
| Constant | 0.477 | | | | | |
| | (0.021) | | | | | |
| Advisor connection | 1.16 | 0.602 | 0.601 | 0.607 | 0.653 | 0.607 |
| | (0.101) | (0.036) | (0.035) | (0.035) | (0.037) | (0.035) |
| PhD's connection | 2.89 | 3.33 | 3.32 | 3.32 | 3.39 | 3.33 |
| | (0.177) | (0.153) | (0.150) | (0.150) | (0.156) | (0.153) |
| *Fixed-effects* | | | | | | |
| PhD Class×Affiliation 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 | 5,512,539 | 5,512,539 | 5,512,539 | 5,512,539 | 5,512,539 | 5,512,539 |

*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. Clustered (PhD university×Field×5 year window + potential hiring university) standard-errors in parentheses

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 to match 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 on average connected to over 10 universities through her advisor.

## Post-PhD outcomes of network hires

We now study whether graduates who are 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 those 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 that are placed through connections are more productive than their classmates not placed through connections. Column 1 shows that, for the number of citations, this gap is 31.1 log points.[5] 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 0.21; further controlling for the advisor's citations before graduation lowers the estimate to 0.16 (standard error 0.022). Thus, even after controlling for these predictors of productivity, a substantial gap persists. It shows that 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 higher citations are partly driven by a higher number of papers. The point estimate is 0.066 (standard error 0.01), indicating that less than half the higher number of citations post-PhD is due to a higher number of papers. Thus, graduates placed through connections are more productive both in terms of 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 still be affiliated with their first university 6 years after the PhD

[5]The point estimate translates into $exp(0.286) - 1 \approx 36\%$ higher citations.

(column 6). This result is not driven by dropping out from research: Column 7 shows that there is no difference in the probability of producing any output 6 years after the PhD or later.

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.

*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 | 0.311 | 0.211 | 0.160 | 0.066 | 0.240 | -0.141 | -0.005 | 0.610 |
| | (0.027) | (0.021) | (0.022) | (0.010) | (0.024) | (0.013) | (0.004) | (0.105) |
| PhD's connection | 0.250 | 0.046 | 0.061 | 0.071 | -0.092 | -0.580 | -0.025 | -0.044 |
| | (0.031) | (0.028) | (0.028) | (0.012) | (0.024) | (0.030) | (0.006) | (0.024) |
| | (0.029) | (0.027) | (0.027) | (0.014) | (0.025) | (0.031) | (0.006) | (0.025) |
| *Fixed-effects* | | | | | | | | |
| PhD Class | 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 $R^2$ | 0.395 | 0.492 | 0.502 | 0.292 | 0.316 | 0.041 | 0.012 | 0.752 |
| Observations | 77,055 | 77,055 | 77,055 | 77,192 | 72,449 | 73,852 | 76,126 | 77,008 |

*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 | 0.154 | 0.072 | 0.0008 | 0.010 | 0.035 | -0.063 | -0.008 |
| | (0.023) | (0.019) | (0.019) | (0.009) | (0.017) | (0.011) | (0.004) |
| PhD's connection | 0.227 | 0.044 | 0.067 | 0.061 | -0.075 | -0.553 | -0.021 |
| | (0.027) | (0.024) | (0.024) | (0.011) | (0.021) | (0.030) | (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 |
| Pre-Graduation Productivity×Field | | Yes | Yes | Yes | Yes | Yes | Yes |
| Advisor Citation Decile×Field | | | Yes | Yes | Yes | Yes | Yes |
| *Fit statistics* | | | | | | | |
| Pseudo $R^2$ | 0.368 | 0.462 | 0.475 | 0.285 | 0.314 | 0.047 | 0.012 |
| Observations | 76,967 | 76,967 | 76,967 | 77,192 | 70,047 | 75,172 | 76,498 |

*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 | 0.120 | 0.060 | 0.009 | 0.022 | 0.046 | -0.058 | -0.005 |
| | (0.023) | (0.021) | (0.021) | (0.010) | (0.018) | (0.013) | (0.005) |
| PhD's connection | 0.206 | 0.026 | 0.041 | 0.052 | -0.092 | -0.560 | -0.023 |
| | (0.028) | (0.027) | (0.027) | (0.012) | (0.023) | (0.031) | (0.006) |
| *Fixed-effects* | | | | | | | |
| PhD Class | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Hiring University Id×Field | 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 $R^2$ | 0.513 | 0.583 | 0.592 | 0.360 | 0.401 | 0.067 | 0.017 |
| Observations | 76,868 | 76,868 | 76,868 | 77,192 | 67,495 | 72,213 | 75,560 |

*Notes:* Unit of observation is a PhD graduate. See equation (2) 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 17% more productive compared to 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 we did in panel A, and show 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 $\gamma$ drops to 0.0008 (standard error 0.019). 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. This implies that networks do not reveal private information about productivity on either side of the market, but that they allow departments to hire PhD graduates that are expected to be more productive.

The remaining columns in panel B keep the same controls as column 3. They show that connected hires, when compared to other hires at the same university, 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 selection 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, that are 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. However, we find no evidence for such effects. Additionally, we find that connected hires are more likely to move to a different university 6 years after the PhD. To the extent that separations depend on the quality of the match, these results also suggest that the network neither reveals information about match quality 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 both for the hiring university and for 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. While network hiring does allocate graduates to more productive places on average, the evidence suggests their productivity would have been similar in the absence of the network connection if working at the same place.

**Discussion** Our results show that from the point of view of a PhD class, connected hires are positively selected in terms of post-PhD career outcomes—ex-post they are more productive, collaborate more with their new colleagues, and are matched to more productive universities. From the point of view of the hiring university, connected hires are also positively selected—but this selection can be predicted by public information at the time of hiring. Thus, conditional on observable characteristics, the selection mechanism of hiring universities is similarly effective in detecting productive graduates for connected and not connected hires.

The results imply that network hiring is effective from the view of job candidates and of hiring universities. PhD graduates are placed at more productive universities and outperform in terms of productivity compared to class peers who are similarly productive before graduation. Hiring universities attract more productive candidates through the network than without.

Overall, we show that hiring through the network is associated with private benefits, but not with aggregate productivity effects. We also do not find evidence that hiring through the network reveals private information, negative or positive.

### The role of the advisor

In the previous sections, we have compared graduates within a PhD class or within the same hiring university. To further sharpen the results we now establish whether the positive selection of connected hires is driven by differences between or within advisors. This also addresses the concern that estimated network effects are driven by more connected advisors having more productive students.

Figure 1 presents the results from variations of the post-PhD regressions. The dots are point estimates; the error bars 95% confidence intervals. In grey, we reproduce the results from Table 3 with controls for PhD graduate productivity during the PhD and advisor cites.

We first compare these baseline estimates to results—shown in yellow—from a model that replaces the advisor cites with advisor fixed effects. The point estimates are very similar between the models. While the standard errors increase, the results remain significant at the five percent confidence level. Thus for the outcomes we consider, the advisor fixed effects account for similar variation in the outcome as the advisor cites—and that our results are not spuriously driven by more connected advisors having more productive students. Instead, the results show that the same advisor—and associated network—place graduates that are ex-post more productive to universities with high output.

Now we compare these latter results to results where we include again fixed effects for the advisor, but we drop the controls for pre-graduation productivity of the PhD graduate. These results are shown in blue; they are very similar to the results from the previous specification. Thus, after accounting for the identity of the advisor, there is limited additional information in the pre-graduation productivity of the PhD graduate in explaining the connected hire gap in post-PhD outcomes.

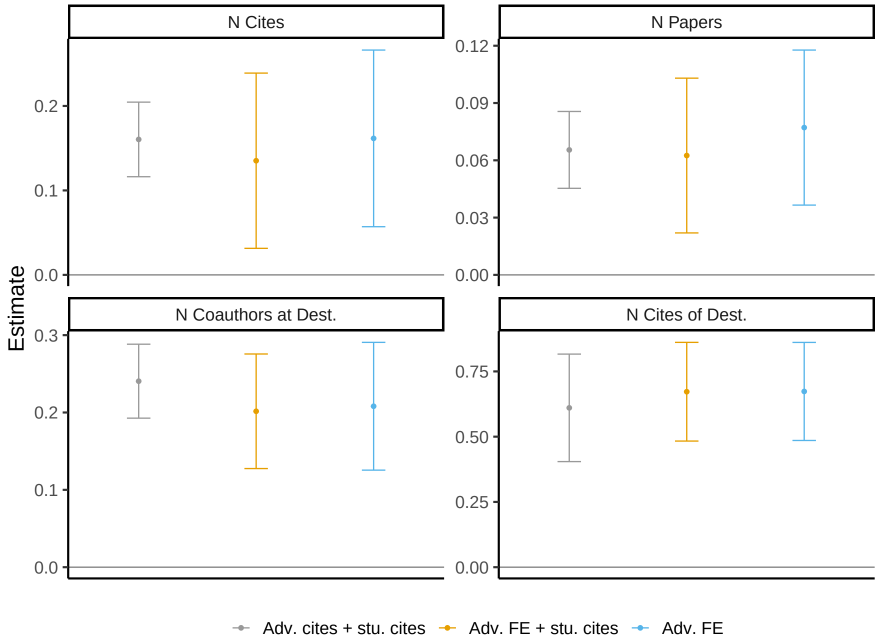


Figure 1: Advisor effect and Post-PhD outcomes

Estimates and associated 95 % confidence intervals of $\gamma$ for the advisor's network. Different colors refer to different specifications of equation (2): *Adv. cites + stu. cites* includes controls for the advisor's and the student's pre-PhD citations; *Adv. FE + stu. cites* includes advisor fixed effects and student pre-PhD citations; and *Adv. FE* includes advisor fixed effects only. Standard errors are clustered at the level of (First Affiliation-Field-5 Year Window).

## Conclusion

We analyze the role of the PhD advisor's collaboration network for the matching of PhD graduates to universities. To do so we build a 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 level of the sending $\times$ hiring university pair, and for the similarity in research topics between the student's dissertation and her potential new colleagues at the destination university.

We then explore whether, on a range of post-hiring performance measures, there is a gap between connected and non-connected graduates. 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 about 15% more citations. When comparing hires at the same university, connected hires are more productive, but 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.

## 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.

**Caplow, Theodore, and Reece J. McGee.** 1958. *The academic marketplace.* Basic Books.

**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.*

**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.

**Hargens, Lowell L, and Warren O Hagstrom.** 1967. “Sponsored and contest mobility of American academic scientists.” *Sociology of education*, 24–38.

**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.

**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.