

# Networks and the market for researchers

Flavio Hafner<sup>1</sup>, Christoph Hedtrich<sup>2</sup>

NBER Summer Institute - Science of Science Funding  
19th of July 2024

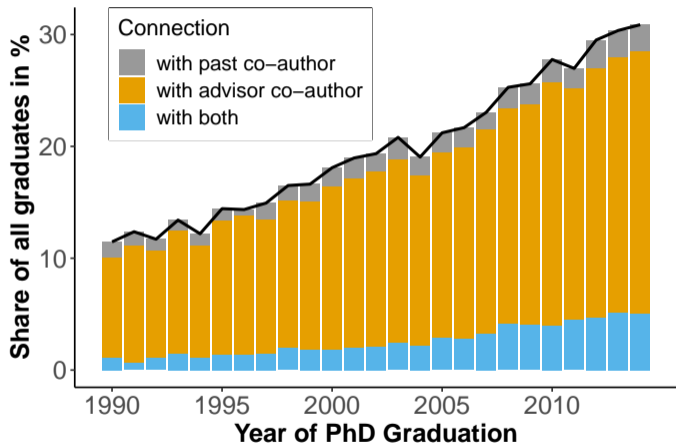
---

<sup>1</sup>Netherlands eScience Center, [f.hafner@esciencecenter.nl](mailto:f.hafner@esciencecenter.nl)

<sup>2</sup>Uppsala University, [christoph.hedtrich@nek.uu.se](mailto:christoph.hedtrich@nek.uu.se)

# Networks and the academic job market

- Research positions require hard to observe skills
- PhD advisors can provide information relevant for matching PhD graduates to universities
- Concerns about favoritism
- Teams are growing → more co-author connections (Fortunato et al., 2018, Freeman et al., 2014, Jones, 2009, Wuchty et al., 2007)



## How are collaboration networks used for placement of PhD graduates?

1. Are co-author connections of advisors predictive of first post-PhD affiliation?
  - ✓  $\approx$  double probability of matching when PhD advisor has a co-author at hiring university
  - ✓ Overall importance of collaboration network for placement doubles due to rise in number of connections from 1990 to 2015

## How are collaboration networks used for placement of PhD graduates?

1. Are co-author connections of advisors predictive of first post-PhD affiliation?
  - ✓  $\approx$  double probability of matching when PhD advisor has a co-author at hiring university
  - ✓ Overall importance of collaboration network for placement doubles due to rise in number of connections from 1990 to 2015
2. Are network hires more productive in first 7 years post PhD?
  - ✓ Productivity premium of 12-30% for network hires within PhD class/hiring university

## How are collaboration networks used for placement of PhD graduates?

1. Are co-author connections of advisors predictive of first post-PhD affiliation?
    - ✓  $\approx$  double probability of matching when PhD advisor has a co-author at hiring university
    - ✓ Overall importance of collaboration network for placement doubles due to rise in number of connections from 1990 to 2015
  2. Are network hires more productive in first 7 years post PhD?
    - ✓ Productivity premium of 12-30% for network hires within PhD class/hiring university
  3. Does productivity premium of network hires reflect private information about match/candidates?
    - ✗ within hiring university: productivity premium predictable by observable characteristics at time of PhD graduation
    - PhD advisors do not reveal private information relevant for productivity
- Co-authorship network of PhD advisors important for *where* PhD graduates end up
- No evidence for direct productivity benefit from network hiring

# Data sources and linking

# Measurement: Research careers of PhD graduates and their advisors

1. Large repository of [PhD theses](#): ProQuest Dissertations & Theses (Clarivate) [PQDT overview](#)
  - basic metadata: year of graduation, advisor
  - content info: title, abstract, subset with full text
2. Bibliometric data on [research output](#): Microsoft Academic Graph (MAG)<sup>3</sup> [MAG overview](#)
  - close to universe of research articles
  - basic metadata (year, authors, affiliations) and citation links

## Main data preparation

- Record linking algorithm with hand-labelled links: [dedupe.io](#)<sup>4</sup> [Linking](#)
- Use MAG natural language model to predict [topics of dissertations](#) from paper abstract

---

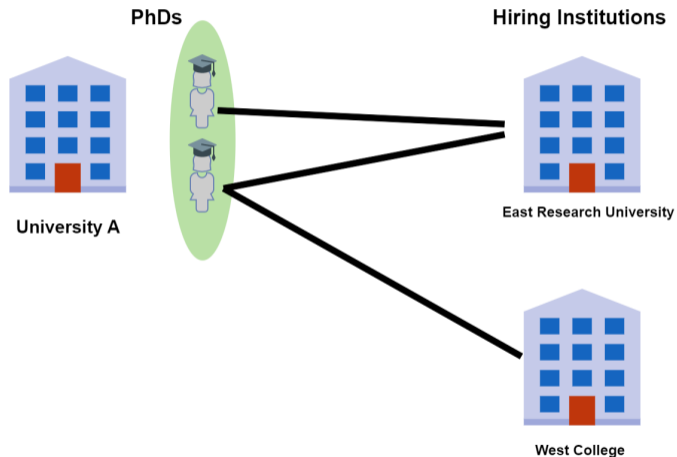
<sup>3</sup>See Wang et al. (2019) for an overview of the data and tools provided.

<sup>4</sup>Python implementation available at <https://github.com/dedupeio/dedupe>

# Empirical Strategy: Networks and Hiring

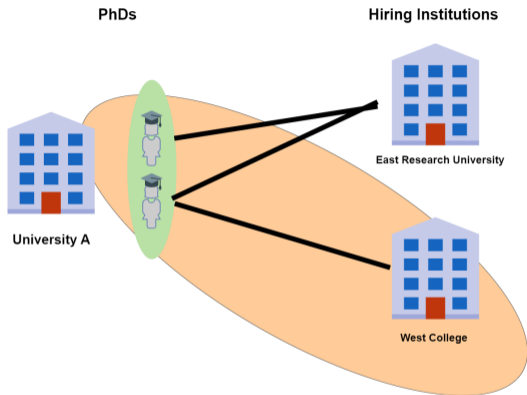


# Network connections are not randomly allocated



- Compare probability to end up at connected university *within* student
- Network connections determined by
  1. geographic distance
  2. subfield specialization
  3. ...
- Same factors determine hiring probability

# Compare graduates only within PhD Class-Hiring University pair



- Compare probability to end up at connected university *within* student
- Control for all systematic factors at the PhD class-hiring university level
- Control for research topic fit between PhD graduate and potential hiring university

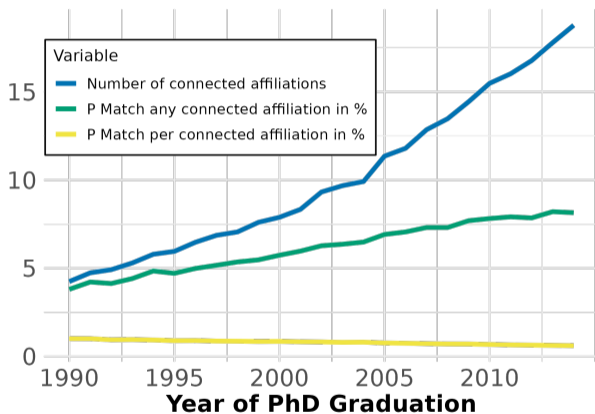
$$\underbrace{E_{i,j}}_{P \text{ } i \text{ matches } j} = \underbrace{\alpha_i}_{\text{student FE}} + \underbrace{\beta_{c(i),j}}_{\text{class-destination effect}} + \underbrace{\gamma_{A,i,j}}_{\text{network effect}} + \underbrace{\delta X_{ij}}_{\text{individual-destination controls}} + \varepsilon_{i,j} \quad (1)$$

# First affiliation of PhD graduates and their advisors co-author network

- Presence of PhD advisor's co-author **doubles** probability of PhD graduate to end up at co-authors university

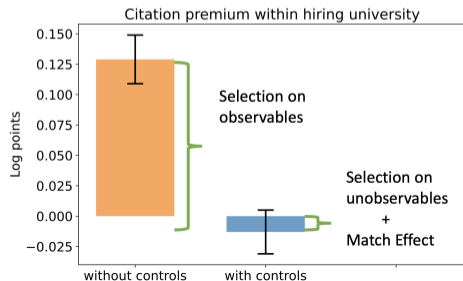
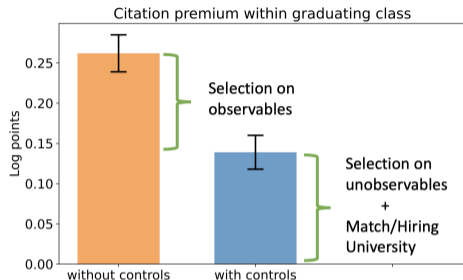
[▶ Table](#)

- Number of connected affiliations triples 1990-2014
- Estimated effect of *one* connection declines
- Total "placement effect" of network connections doubles 1990-2014



**Are network hires more productive?**

# Network hires are more productive, predictably so [▶ Specification](#)



- PhD graduates hired via network are more productive
    - compared to PhD classmates [▶ Table](#)
    - within hiring university [▶ Table](#)
  - Hiring universities receive "better" candidates via network
  - Productivity premium of network hires is predictable by
    - student's publications during PhD
    - prestige of PhD advisor
- No evidence for selection on unobserved (match) quality of network hires

## Conclusion

- We observe large share of labor market of early career scientists & occupation specific productivity measures, both pre- and post-hiring
- Collaboration network of PhD advisor → hiring network for PhD graduate
- Importance of co-author connections for placement doubles between 1990 and 2014
- Comparing post PhD outcomes we find:
  - network hires positively selected
  - no evidence that the collaboration network reveals private information about PhD graduates or the particular match with a university
- Networks widely used for hiring, but no evidence for direct productivity benefit from network hiring

# First affiliation of PhD graduates and co-author networks

[Back](#)
[by advisor prominence](#)
[Sorting](#)
[Gender gap](#)

Dependent Variable:	Match formed					
Model:	(1)	(2)	(3)	(4)	(5)	(6)
<i>Variables</i>						
Constant	0.477 (0.021)					
Advisor connection	1.16 (0.101)	0.602 (0.036)	0.601 (0.035)	0.607 (0.035)	0.653 (0.037)	0.607 (0.035)
PhD's connection	2.89 (0.177)	3.33 (0.153)	3.32 (0.150)	3.32 (0.150)	3.39 (0.156)	3.33 (0.153)
<i>Fixed-effects</i>						
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
<i>Additional controls with varying slopes</i>						
Max similarity to faculty members×Field						Yes
Avg. similarity to faculty members×Field						Yes
Observations	5,511,980	5,511,980	5,511,980	5,511,980	5,511,980	5,511,980

## Are network hires more productive? [▶ Back](#)

- Output  $y_i$ : research output in 7 years after PhD graduation
- $A_{i,j(i)}$ : First affiliation  $j(i)$  had a network connection

$$y_i = \exp \left( \underbrace{\alpha_{c(i)}}_{\text{PhD class FE}} + \underbrace{\alpha_{j(i)}}_{\text{First affil FE}} + \underbrace{\delta X_i}_{\text{pre-graduation controls}} + \underbrace{\gamma A_{i,j(i)}}_{\text{connected hire}} + u_i \right), \quad (2)$$

where

- $c(i)$  is the class of graduate  $i$ ,  $j(i)$  the first affiliation
- $\alpha_{c(i)}$ ,  $\alpha_{j(i)}$  captures PhD class/first affiliation attributes
- Compare graduates within class, and within hiring university
- $X_i$  controls for graduate  $i$ : pre graduation citation deciles of graduate and advisor
- $\gamma$ : output difference between graduates with a first affiliation with a network connection vs. without a connection

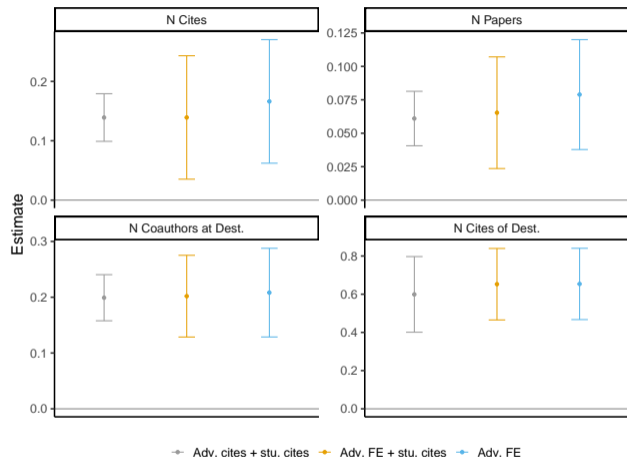


# PhD advisor network and outcomes of graduates (i) - Within Class

[▶ Back](#)

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)
<i>Variables</i>								
Advisor connection	0.262 (0.023)	0.186 (0.019)	0.139 (0.021)	0.061 (0.010)	0.199 (0.021)	-0.143 (0.013)	-0.005 (0.004)	0.599 (0.101)
PhD's connection	0.215 (0.029)	0.034 (0.027)	0.048 (0.027)	0.069 (0.012)	-0.086 (0.022)	-0.588 (0.029)	-0.024 (0.006)	-0.028 (0.025)
<i>Fixed-effects</i>								
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
<i>Fit statistics</i>								
Pseudo R <sup>2</sup>	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

## PhD advisor network and outcomes of graduates (ii) - Within Advisor [▶ Back](#)



- *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
- *Adv. FE* includes advisor fixed effects only.

Figure: Advisor and Post-PhD outcomes

## Within Class comparison of Post-PhD outcomes [▶ Back](#)

- Network hires are more productive than their classmates:  $> 30\%$  more citations on papers in first 7 years post-PhD
- $\approx 50\%$  productivity premium is explained by pre-graduation productivity of graduate *and* advisor
- Network hires are placed into more productive environments, and collaborate more with their new colleagues
- Productivity differences reflect *both* selection of productive candidates and effect of placement

# PhD advisor network and outcomes of graduates (iii) - Within Hiring University

[▶ Back](#)

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)
<i>Variables</i>							
Advisor connection	0.129 (0.020)	0.058 (0.018)	-0.013 (0.018)	0.006 (0.009)	0.026 (0.015)	-0.065 (0.012)	-0.009 (0.005)
PhD's connection	0.191 (0.026)	0.026 (0.024)	0.048 (0.024)	0.059 (0.011)	-0.077 (0.018)	-0.559 (0.029)	-0.022 (0.006)
<i>Fixed-effects</i>							
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
<i>Fit statistics</i>							
Pseudo R <sup>2</sup>	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

## Within Hiring University comparison of Post-PhD outcomes [▶ Back](#)

- Network hires are more productive than non-connected hires at same university:  $> 16\%$  more citations on papers in first 7 years post-PhD
- Network hires positively selected
- Productivity premium is explained by observable pre-graduation productivity of graduate *and* advisor
- No revelation of unobserved information about productivity through network
- Results consistent with lower hiring costs through network

# PhD advisor network and outcomes of graduates (iv) - PhD Class and Hiring University Fixed Effects

Dependent Variables:	N Cites PhD graduate			N papers	Co-authors	Same Affil	Any output
	(1)	(2)	(3)	(4)	First Affil	PhD+6yrs	PhD+6yrs
<i>Variables</i>							
Advisor connection	0.091 (0.021)	0.041 (0.020)	-0.008 (0.019)	0.018 (0.010)	0.035 (0.018)	-0.059 (0.013)	-0.006 (0.005)
PhD's connection	0.174 (0.028)	0.012 (0.027)	0.026 (0.027)	0.050 (0.013)	-0.090 (0.022)	-0.571 (0.030)	-0.022 (0.006)
<i>Fixed-effects</i>							
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
<i>Fit statistics</i>							
Pseudo R <sup>2</sup>	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

- Results almost identical to just controlling for Hiring University FE
  - PhD class does not have much additional information about productivity premium of connected hires
- Selection mechanism of universities on average reveals information about graduates future productivity similarly for connected and non-connected hires

## Post-PhD outcomes: Assessing Changes over time

- We test for time trends using this specification:

$$y_{i,j(i)} = \exp \left( \alpha_{c(i)} + \alpha_{j(i)} + \beta X_i + \sum_F \gamma_F A_{i,j} \mathbf{1}\{F = \text{Field}_i\} + \underbrace{\delta A_{i,j} \times (t(i) - 1990)}_{\text{linear time trend}} + u_i \right), \quad (3)$$

- estimates for  $\delta$  are small, positive and not statistically significant Results

→ reject large trend towards more negative selection



# Post PhD Outcomes - Time Trend [Back](#)

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)
<i>Variables</i>					
Advisor connection $\times (t - 1990)$	0.002 (0.003)	0.0006 (0.001)	0.0007 (0.002)	0.001 (0.002)	0.0008 (0.0006)
PhD's connection $\times (t - 1990)$	0.009 (0.005)	0.001 (0.002)	0.002 (0.004)	0.005 (0.004)	0.003 (0.0009)
<i>Fixed-effects</i>					
PhD Class	Yes	Yes	Yes	Yes	Yes
Pre Graduation Productivity $\times$ Field	Yes	Yes	Yes	Yes	Yes
Advisor Citation Decile $\times$ Field	Yes	Yes	Yes	Yes	Yes
Hiring University Id $\times$ Field	Yes	Yes	Yes	Yes	Yes
Subfield (MAG lvl 1)	Yes	Yes	Yes	Yes	Yes
Degree Year $\times$ Field	Yes	Yes	Yes	Yes	Yes
Advisor connection $\times$ Field	Yes	Yes	Yes	Yes	Yes
PhD's connection $\times$ Field	Yes	Yes	Yes	Yes	Yes
<i>Fit statistics</i>					
Observations	73,596	73,885	64,625	70,109	72,893
Pseudo R <sup>2</sup>	0.62	0.37	0.43	0.07	0.015

# First affiliation of PhD graduates by gender

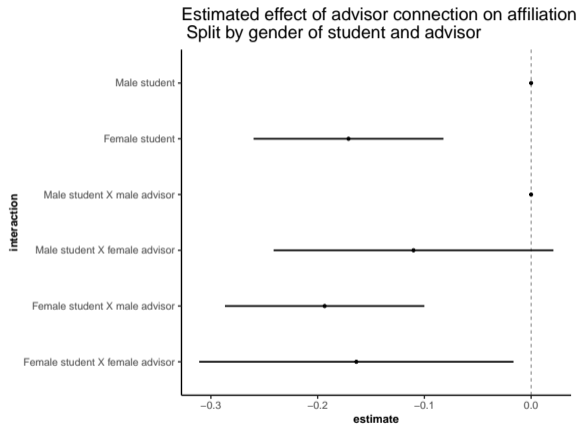


Figure: Overall gender gap

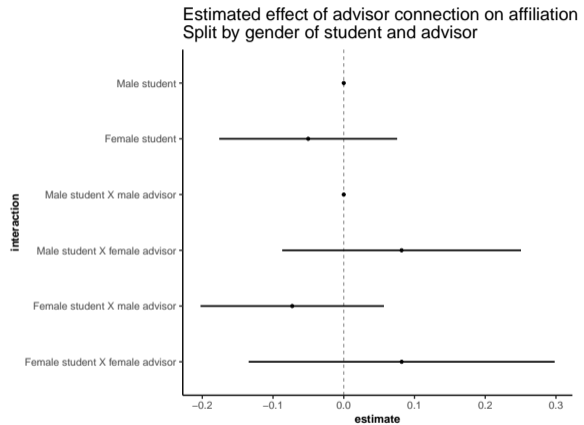


Figure: Within PhD class-potential hiring university gender gap

# First affiliation of PhD graduates by origin and destination

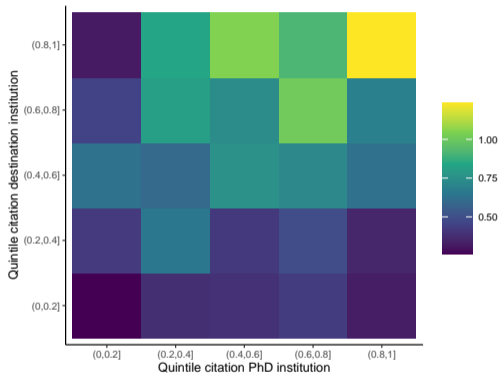


Figure: Network effect  $\gamma$  of advisor connection by Origin and Destination Institution Rank

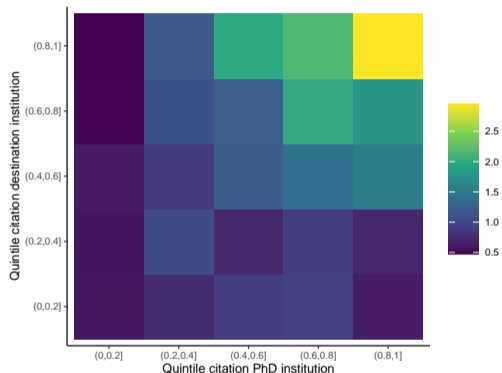


Figure: Total effect including number of connections

## References I

- Fortunato, S., C. T. Bergstrom, K. Börner, J. A. Evans, D. Helbing, S. Milojevi, A. M. Petersen, F. Radicchi, R. Sinatra, B. Uzzi, et al. (2018): “Science of science,” *Science*, 359, eaao0185.
- Freeman, R. B., I. Ganguli, and R. Murciano-Goroff (2014): “Why and wherefore of increased scientific collaboration,” in *The changing frontier: Rethinking science and innovation policy*, University of Chicago Press, 17–48.
- Jones, B. F. (2009): “The burden of knowledge and the “death of the renaissance man”: Is innovation getting harder?” *The Review of Economic Studies*, 76, 283–317.
- Wang, K., Z. Shen, C. Huang, C.-H. Wu, D. Eide, Y. Dong, J. Qian, A. Kanakia, A. Chen, and R. Rogahn (2019): “A review of microsoft academic services for science of science studies,” *Frontiers in Big Data*, 2, 45.
- Wuchty, S., B. F. Jones, and B. Uzzi (2007): “The increasing dominance of teams in production of knowledge,” *Science*, 316, 1036–1039.