Academic Mobility in Economics

by Juan Carlos Cisneros, Srividya Dasaraju, and Sameer Nair-Desai

April 16, 2024

Summary | TLDR

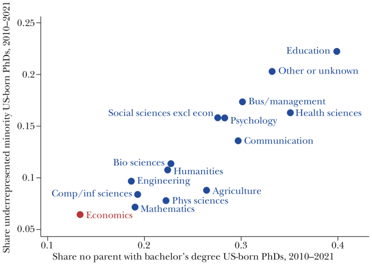
- **Goal:** We plan to build a rich panel dataset of **Economics researchers*** that **charts idiosyncratic characteristics predictive of *academic mobility*.**[1]

- **Research Question:** We intend to **(1)** distinguish the relative importance of distinct factors predicting academic "success"; and **(2)** provide suggestive evidence on the consequences of academic "elitism" for diversity and knowledge generation. Specifically:
    - How do we conceptualize academic "success"? We propose measures incorporating reputation, resource access, influence, and innovation.
    - What is the relative importance of one's initial endowments, PhD advisors or mentors, coauthorship ties, and institutions in predicting academic "success"? How does this predictive power vary over time and space?
        - Which network ties are most impactful for career advancement? Which academics serve as "bridges" for the underrepresented?
        - Who are the outlier "success stories" in the profession? How can we explain their upward trajectory in the absence of traditional advantages?
    - How do observed inequities in these domains affect scientific production?

[1] We plan to trace scholars from the 1970 Economics PhD cohort until present day.

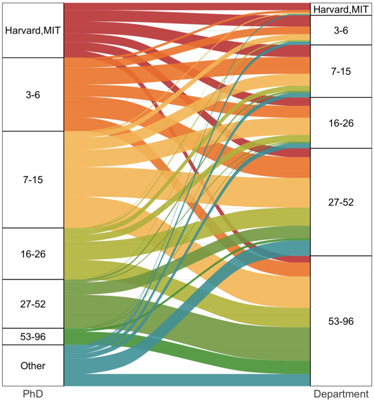
Motivation: (Lack of) Diversity and Elitism in Economics

Panel B. Underrepresented minority share and first-generation share



Source: Survey of Earned Doctorates.
Note: Underrepresented minorities include racial or ethnic minorities.

Figure 8: Flows From PhD Programs (Left) to Departments (Right), by Category



Notes: This Sankey diagram shows flows from PhD programs (left) to departments (right). The width of the flow represents the number of individuals going from one group to another.

- Figure from: "The Economics Profession's Socioeconomic Diversity Problem".
- Figure from: "Staying at the Top: The Ph.D. Origins of Economics Faculty".

Possible explanations from the literature

- Emphasis on "elite" journals/institutions incentivizes careerism and clientele effects (Heckman and Moktan, 2020).
- Disparate access to professional relationships and academic networks worsens inequality (Colussi, 2018; Stansbury and Schultz, 2023).
- High costs and information asymmetries in graduate education pipelines reduces choice set (Bayer and Rouse, 2016; Stansbury and Schultz, 2023).
- Lack of role models for those underrepresented (Porter and Serra, 2020).
- Institutional de facto barriers in publishing/hiring/promotion stages (Iaria et al., 2024; Card et al., 2022, 2020; Hengel, 2022; Zacchia, 2021).
- Journals/events/awards with high reputation are highly centralized in only elite institutions in the US and Europe (Amarante et al., 2022).

What Shapes an Academic Trajectory?

$t = 0$ **(point of entry: JMP)**

- Academic Institution (undergrad/grad)
- PhD advisors, mentors, peers
- Coauthorships
- Publications and citations
- Demographics and initial endowments (i.e., academic parent)
- Funding
- Field
- RAships or predoc experiences

$t = T$ **(end of career; or 2024)**

- Cumulative Publications (Top 5s), impact factor, citations
- Full coauthorship network
- Awards, distinctions
- Academic rank and tenure status
- Total funding
- Editorial roles or service positions

Predictive Inference

To evaluate the predictive power and stability of potential variables that explain the variation of different measures of academic success, we intend to:

- Following Chetty et al. (2022): Measure incremental R^2 of each covariate on a regression that includes the other explanatory variables and measure the weights placed on each variable by a Lasso regression.
- Select the best-performing ensemble learning algorithm based on out-of-sample MSE and evaluate feature importance with Shapley values.
- Check the stability of predictors across time and subgroups:
  - Validate cross-sectional fit of model across years.
  - Estimate heterogeneous predictive effects.

Data

- ProQuest Dissertations & Theses Global
  - From this data, we can identify economists from all graduating cohorts from 2007 onwards, and parse their job market paper (JMP). The JMP has information on their advisor, members of their committee, and field.
  - This enables us to see scholars who are influential today at the time they "entered" the discipline, *and* to trace those who did not end up in academia.
- Publications Dataset
  - Collates: {Web of Science, EconLit, JSTOR Constellate, SCOPUS}.
- Citations Dataset
  - Collates: {CiTeC, Google Scholar}.
- Academic Network Graphs
  - Collates: {RePeC, OpenAlex, Dimensions}.
- Researcher Demographics
  - Scraped from online or pulled from existing lit. (i.e., Card et al. (2022)).
- Academic CVs and Professional Webpages

Dataset structure

- Table 1 visualizes our **author-year level** dataset.

Table: Author-Year Level

Researcher	Year	Tenure Status	Academic Rank	Citations	RePeC Index	Attributes
Name 1	2014	Yes	Full Prof.	12034	103	X_i
Name 1	2013	Yes	Associate Prof.	11032	124	X_i
Name 2	2022	No	Asst. Prof.	1232	2421	X_i
...	...	...	...	...	...	...
...						

- Table 2 contains the **relationships** between the researchers in a given year.

Table: Coauthorship Edges

$Researcher_i$	$Researcher_j$	Link type	Year
Name 1	Name 2	Coauthor	2014
Name 1	Name 3	$Author_i - RA_j$	2014
...	...		

Illustration

Use automated data extraction (LabelStudio + document mining) to create new author-level network features:

Figure: Mentor-mentee link

It is a real pleasure to thank the many people who have supported me and my research throughout my years at Northwestern. First and foremost, I am extremely indebted to my advisor and mentor Joel Mokyr, who intrigued my interest in economic history and encouraged this project from its very first stage. Joel's continuous guidance, energy and contagious enthusiasm turned my Ph.D into a pleasant experience. Joel read carefully and commented thoughtfully on every single word I have written and determinedly challenged me to improve the quality of my work.

Illustration

Figure: Acknowledgements

David Card: card@berkeley.edu
Stefano DellaVigna: sdellavi@econ.berkeley.edu
Patricia Funk: patricia.funk@usi.ch
Nagore Iriberri: nagore.iriberri@gmail.com

We thank the Executive Committee of the Econometric Society for pursuing this study and for agreeing to share the confidential data on nominations for Econometric Society Fellows. We thank Vincent P. Crawford, Avinash Dixit, Robert Gordon, Dan Hamermesh, Erin Hengel, Shelly Lundberg, Enrico Moretti, Muriel Niederle, Rafael Repullo, Heather Sarsons, Joel Sobel, and Heidi Williams and seminar audiences at the COSME conference, at Stanford University, UCL, the University of Bolzano, the University of Vienna, and Washington University for comments and Robert Gordon, Dan Hamermesh, Rafael Repullo, Jose Scheinkman, and Chris Shannon for helping with various aspects of the data collection. We are grateful to Joshua Archer, Zihao Li, Chris Lim, Dan Ma, Sergio Nascimento, Anna Sun, Henry Xu, and a team of undergraduate research assistants for their extraordinary help. Nagore Iriberri acknowledges financial support from Grants PID2019-106146GB-I00 MINECO/FEDER and IT367-19. Patricia Funk acknowledges financial support from the Swiss National Science Foundation (Grant 178887).

Figure: Acknowledgements - Labelled

Network Analysis: Setup

- From the source datasets: we build the series of graphs $\{G\}_{t=min(year)}^{max(year)} = \{V_t, E_t, \Lambda_t, \Omega_t\}_{t=min(year)}^{max(year)}$, where:
  - V_t are the nodes of the graph
  - E_t are the edges of the graph
  - Λ_t is a matrix of node attributes
  - Ω_t is a vector of edge attributes
- In the final dataset we map the information on network positions, roles, and dynamics in both networks to a researcher-year panel. For that purpose, we introduce the concept of a cumulative network:

Definition

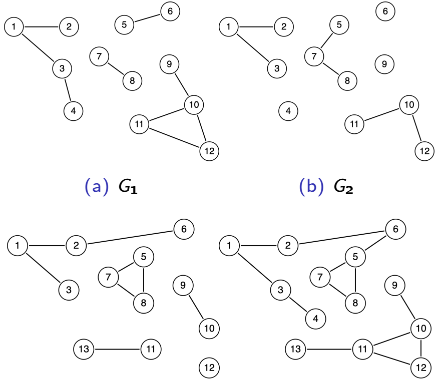
Cumulative Network ($\tilde{G}_T$):

$\tilde{G}_T = [\{V_t\}_{min(t)}^{T}, \{E_t\}_{min(year)}^{T}, f(\{\Lambda_t\}_{min(t)}^{T}), g(\{\Omega_t\}_{min(t)}^{T})]$, where $T \geq t$, includes all the nodes and edges from all the previous periods and aggregates the information from all Λ_t and Ω_t using real-valued functions f($\cdot$) and g($\cdot$).

Network Analysis: Illustration

A cumulative network allows both direct and network effects (i.e., it can incorporate the attributes of “neighbor" nodes, the attributes of their k-hop neighborhoods, and the attributes of the full network).

Figure: Cumulative network illustration



(a) G_1 (b) G_2

(c) G_3 (d) $\tilde{G}_3$

Network Analysis: Recipes

Query: What is the weight of one's initial coauthors in their long-term success?

Recipe 1: Interpretable variables

1. Map features of $\tilde{G}_T$ to node and edge attributes Λ_1 and Ω_1. This can be done parametrically (e.g., using a discount factor) and/or through network metrics. We denote this graph as the present value network G_1^{PV}.
2. Map features of G_1^{PV} to an author-level data frame as network metrics, co-author level variables, and/or neighborhood level variables.
3. Include these new variables in the predictive inference models.

Recipe 2: High-dimensional embeddings (Chernozhukov et al., 2024)

1. Embed the nodes in $\tilde{G}^T$ using representation learning algorithms (e.g., train a transformer-based architecture using edges as the ground truth).
2. Construct an author-level data frame with $t = 0$ covariates X_{i0} and a measure of academic success y_i. Regress $X_{0,i}$ on y_i using OLS, Lasso, and any machine learning model and measure the covariates' incremental R^2.
3. Partial out the embedding interacted with E_1 from both X_{i0} and y_i using the flexible machine learning model with best out of sample performance; denote the residualized variables $\tilde{X}_{0i}$ and $\tilde{y}_i$. Repeat this using $\tilde{X}_{i0}$ and $\tilde{y}_i$.

References I

Amarante, V., Burger, R., Chelwa, G., Cockburn, J., Kassouf, A., McKay, A., and Zurbrigg, J. (2022). Underrepresentation of developing country researchers in development research. *Applied Economics Letters*, 29(17):1659–1664.

Bayer, A. and Rouse, C. E. (2016). Diversity in the economics profession: A new attack on an old problem. *Journal of Economic Perspectives*, 30(4):221–242.

Card, D., DellaVigna, S., Funk, P., and Iriberri, N. (2020). Are referees and editors in economics gender neutral? *The Quarterly Journal of Economics*, 135(1):269–327.

Card, D., DellaVigna, S., Funk, P., and Iriberri, N. (2022). Gender differences in peer recognition by economists. *Econometrica*, 90(5):1937–1971.

Chernozhukov, V., Hansen, C., Kallus, N., Spindler, M., and Syrgkanis, V. (2024). Applied causal inference powered by ml and ai. *rem*, 12(1):338.

Chetty, R., Jackson, M. O., Kuchler, T., Stroebel, J., Hendren, N., Fluegge, R. B., Gong, S., Gonzalez, F., Grondin, A., Jacob, M., et al. (2022). Social capital i: measurement and associations with economic mobility. *Nature*, 608(7921):108–121.

References II

Colussi, T. (2018). Social ties in academia: A friend is a treasure. *Review of Economics and Statistics*, 100(1):45–50.

Heckman, J. J. and Moktan, S. (2020). Publishing and promotion in economics: The tyranny of the top five. *Journal of Economic Literature*, 58(2):419–470.

Hengel, E. (2022). Publishing while female: Are women held to higher standards? evidence from peer review. *The Economic Journal*, 132(648):2951–2991.

Iaria, A., Schwarz, C., and Waldinger, F. (2024). Gender gaps in academia: Global evidence over the twentieth century. *Available at SSRN 4150221*.

Porter, C. and Serra, D. (2020). Gender differences in the choice of major: The importance of female role models. *American Economic Journal: Applied Economics*, 12(3):226–254.

Stansbury, A. and Schultz, R. (2023). The economics profession's socioeconomic diversity problem. *Journal of Economic Perspectives*, 37(4):207–230.

Zacchia, G. (2021). What does it take to be top women economists? an analysis using rankings in repec. *Review of Political Economy*, 33(2):170–193.