

What drives the productivity of scientific labor?

Aaron Clauset
@aaronclauset
Professor
Computer Science Dept. & BioFrontiers Institute
University of Colorado, Boulder
External Faculty, Santa Fe Institute



network data science



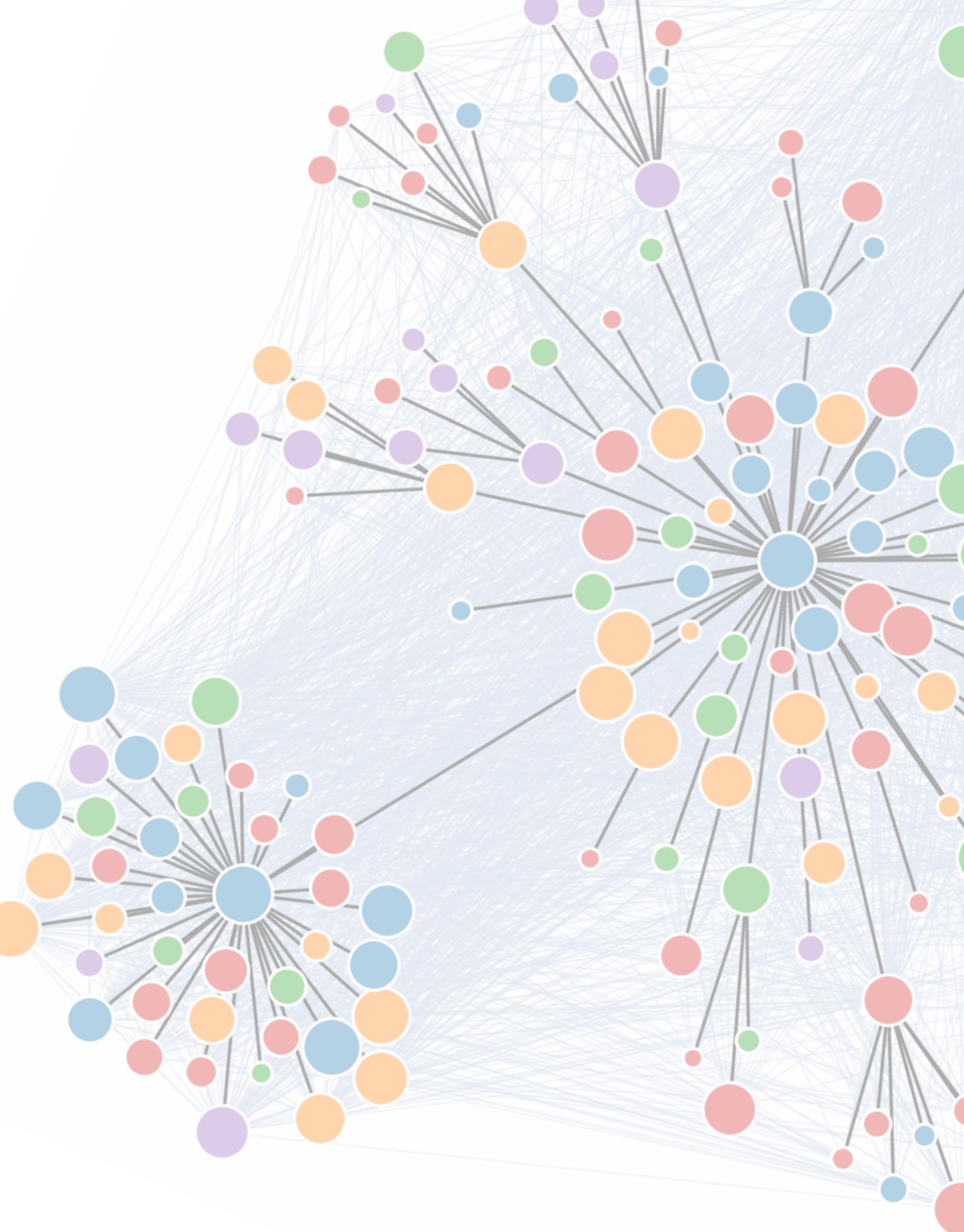
University of Colorado **Boulder**

Network Analysis and Modeling

This graduate-level course will examine modern techniques for analyzing and modeling the structure and dynamics of complex networks. The focus will be on statistical algorithms and methods, and both lectures and assignments will emphasize model interpretability and understanding the processes that generate real data. Applications will be drawn from computational biology and computational social science. No biological or social science training is required. (Note: this is not a scientific computing course, but there will be plenty of computing for science.)

Full lectures notes online (~150 pages in PDF)

<https://aaronclauset.github.io/courses/5352/>



icon.colorado.edu

ICON

Colorado Index of Complex Networks

A comprehensive index of research-quality network data sets

What is ICON?

The Colorado Index of Complex Networks (ICON) is a comprehensive index of research-quality network data sets from all domains of network science, including social, web, information, biological, ecological, connectome, transportation, and technological networks.

Each network record in the index is annotated with and searchable or browsable by its graph properties, description, size, etc., and many records include links to multiple networks. The contents of ICON are curated by volunteer experts from Prof. Aaron Clauset's research group at the University of Colorado Boulder.

Click [NETWORKS](#) to view the networks in the index.

Network editors & packages

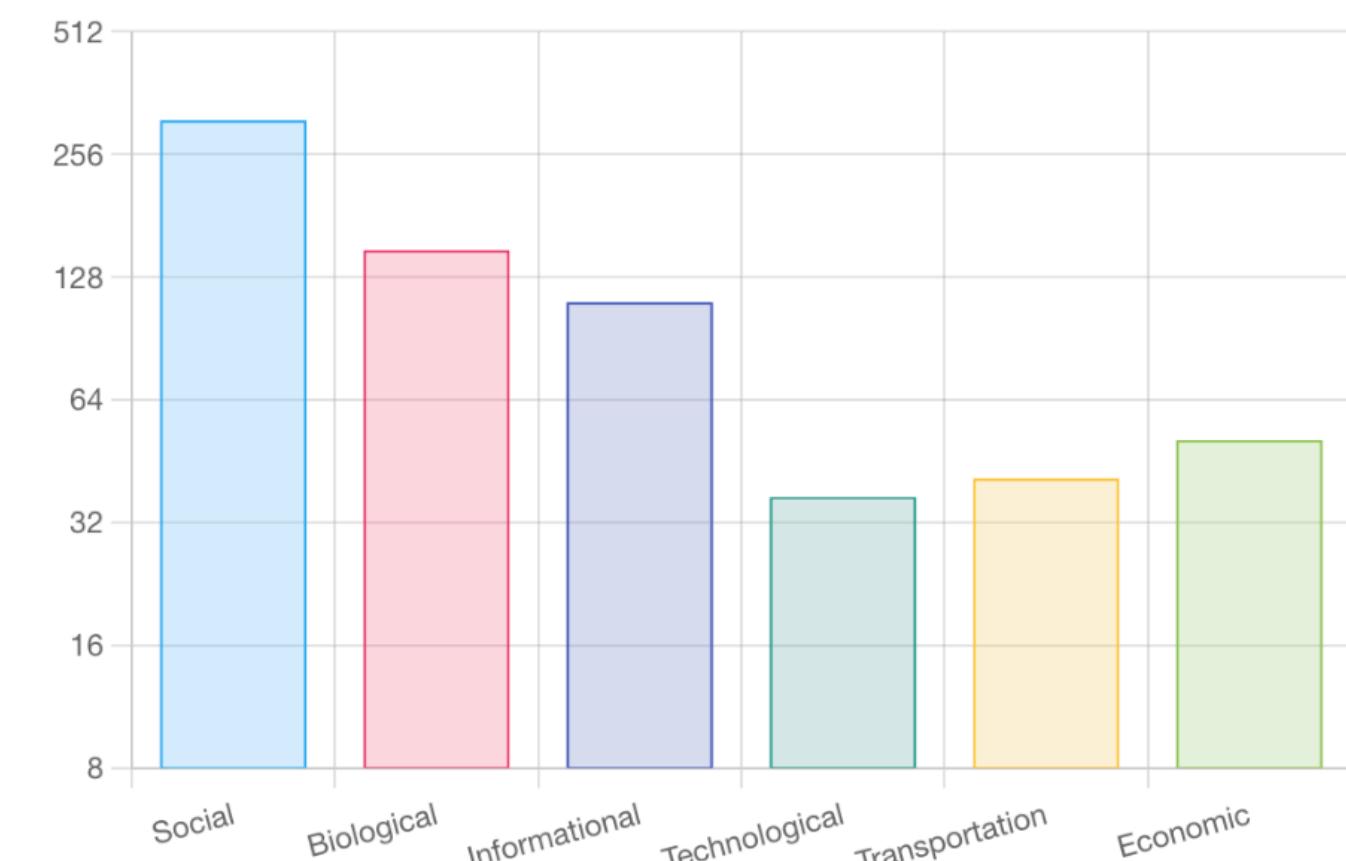
[NetworkX](#) [python]
[igraph](#) [python, R, c++]
[graph-tool](#) [python, c++]
[GraphLab](#) [python, c++]

[UCI-Net](#)
[NodeXL](#)
[Gephi](#)
[Pajek](#)
[Network Workbench](#)
[Cytoscape](#)
[yEd graph editor](#)
[Graphviz](#)

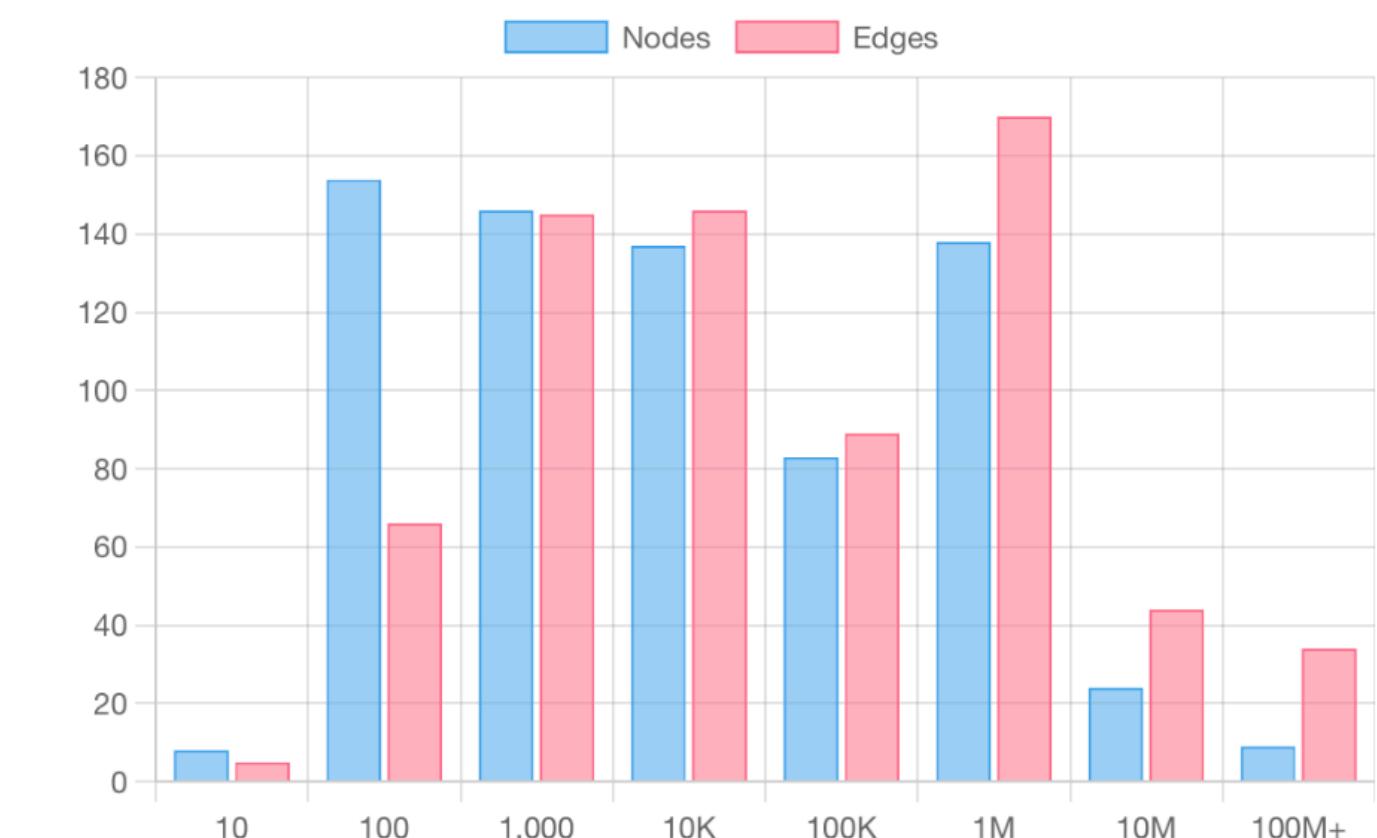
Network data sets

[Colorado Index of Complex Networks](#)
[icon.colorado.edu](#)

Network Domains



Network Nodes and Edges Distribution



science is not a meritocracy

science is not a meritocracy

good ideas do not always win out over mediocre ones



poor experimental design & causal misinterpretations

The screenshot shows a news article from Scientific American. The title is "The D...". Below it, a sub-headline reads "Vibrational spectra in...". The text "The least skilled people..." is visible. At the bottom, it says "Science 377, (2010)" and "Scientific American".

Power Posing: Brief Nonverbal Displays Affect Neuroendocrine Levels and Risk Tolerance

Psychological Science 21 (2010)

Dana R. Carney¹, Amy J.C. Cuddy², and Andy J. Yap¹

¹Columbia University and ²Harvard University

[View Article Online](#) | [Download PDF](#) | [Email this article](#)

Female hurricanes are deadlier than male hurricanes

Kiju Jung^{a,1}, Sharon Shavitt^{a,b,1}, Madhu Viswanathan^{a,c}, and Joseph M. Hilbe^d

PNAS 111 (2014)

science is not a meritocracy

good ideas do not always win out over mediocre ones



A screenshot of a news article from Scientific American. The title reads "BLO The D" and the subtitle "An Sce". A large orange box contains the text "these are 'epistemic inefficiencies'".

Power Posing: Brief Nonverbal Displays Affect Neuroendocrine Levels and Risk Tolerance

¹Columbia University and ²Harvard University

[View Article Online](#)

Ellis R. Lippincott, Robert R. Stromberg,
Warren H. C

Science 164 (1969)

Female hurricanes are deadlier than male hurricanes

Kiju Jung^{a,1}, Sharon Shavitt^{a,b,1}, Madhu Viswanathan^{a,c}, and Joseph M. Hilbe^d

PNAS 111 (2014)

science is not a meritocracy

good ideas do not always win out over mediocre ones

the best scientists are not reliably recognized



science is not a meritocracy

good ideas do not always win out over mediocre ones

the best scientists are not reliably recognized



HISTORY OF INNOVATION

Who do we invent for? Patents by women focus more on women's health, but few women get to invent

Rembrand Koning^{1*}, Sampsa Samila², John-Paul Ferguson³

Science 372 (2021)

WHO BECOMES AN INVENTOR IN AMERICA?
THE IMPORTANCE OF EXPOSURE TO INNOVATION*

ALEX BELL
RAJ CHETTY
XAVIER JARAVEL
NEVIANA PETKOVA
JOHN VAN REENEN

The Quarterly J. Economics (2019)

* "Lost Einsteins" paper

science is not a meritocracy

good ideas do not always win out over mediocre ones

the best scientists are not reliably recognized



HISTORY OF INNOVATION

Who do we invent for? Patents by women focus more on women's health, but few women get to invent

Rembrand Koning^{1*}, Sampsa Samila², John-Paul Ferguson³

Science 372 (2021)

The Diversity-Innovation Paradox in Science
SCIENTIFIC COMMUNITY

The unequal impact of parenthood in academia

Allison C. Morgan^{1*}, Samuel F. Way¹, Michael J. D. Hoefer¹, Daniel B. Larremore^{1,2},
Mirta Galesic³, Aaron Clauset^{1,2,3*}

Science Advances 7 (2021)

women experience a negative "shock" to their productivity

WHO BECOMES AN INVENTOR IN AMERICA?
THE IMPORTANCE OF EXPOSURE TO INNOVATION*

ALEX BELL

RAJ CHETTY

XAVIER JARAVEL

JOVA
NEN

Dan Jurafsky^{b,c},

S 117 (2020)

* "Lost Einsteins" paper

science is not a meritocracy

good ideas do not always win out over mediocre ones

the best scientists are not reliably recognized



HISTORY OF INNOVATION

Who do we invent for? Patents by women focus more

WHO BECOMES AN INVENTOR IN AMERICA?

ON*

social biases

drive epistemic inefficiencies

SCIENTIFIC COMMUNITY

The unequal impact of parenthood in academia

Allison C. Morgan^{1*}, Samuel F. Way¹, Michael J. D. Hoefer¹, Daniel B. Larremore^{1,2},
Mirta Galesic³, Aaron Clauset^{1,2,3*}

NEN
Dan Jurafsky^{b,c},

S 117 (2020)

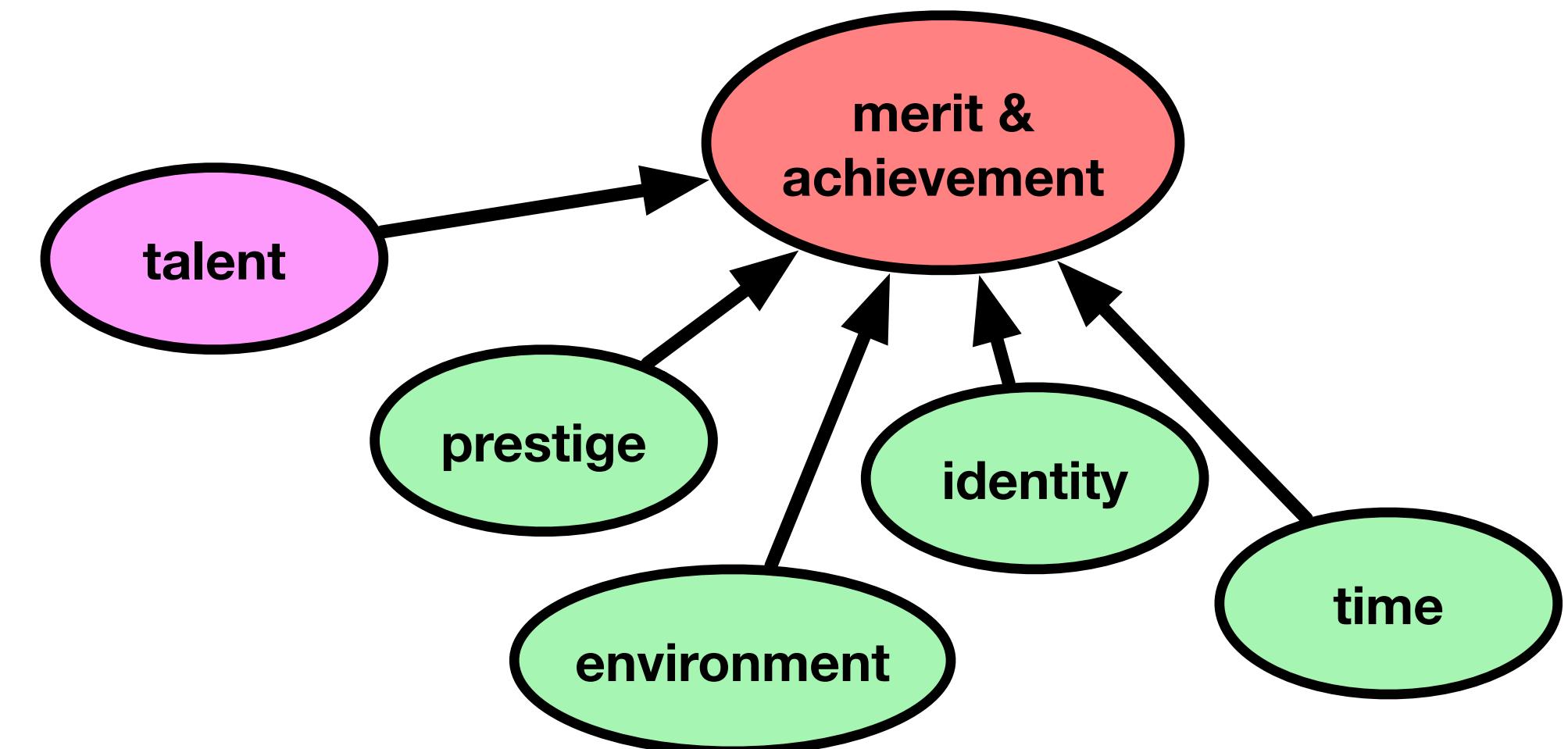
* "Lost Einsteins" paper

science is not a meritocracy

good ideas do not always win out over mediocre ones

the best scientists are not reliably recognized

"merit" is entangled with other factors like **prestige, environment, identity, and time**



science is not a meritocracy

good ideas do not always win out over mediocre ones

the best scientists are not reliably recognized

"merit" is entangled with other factors like **prestige, environment, identity, and time**

- ▶ cannot *fairly* compare two scientists directly
 - if they differ in their environment, time, etc.
 - ignoring non-meritocratic effects overvalues 'winners' (Matthew Effect)

science is not a meritocracy

good ideas do not always win out over mediocre ones

the best scientists are not reliably recognized

"merit" is entangled with other factors like **prestige, environment, identity, and time**

- ▶ cannot *fairly* compare two scientists directly
 - if they differ in their environment, time, etc.
 - ignoring non-meritocratic effects overvalues 'winners' (Matthew Effect)

if it's not a level playing field, what is it?

science is an ecosystem



not a new idea! many others (including us) argue this
useful as a metaphor
but also as a *model*

The image shows the front cover of the February 10, 2017, issue of the journal *Science*. The cover features a large, glowing crystal ball in the foreground, with the word "PREDICTION" written across it in white. Above the crystal ball, the word "SCIENCE" is partially visible. The main title "PREDICTION" is prominently displayed in large, bold letters. Below the title, the subtitle "What can we know in advance about human activities?" is written. At the top of the cover, there are several small text boxes with headlines: "The science behind a \$2 billion biotech gamble", "Deep-ocean monitoring for climate protection", and "Synapses shrink during sleep, so do we". The journal's name "Science" and its publisher "AAAS" are at the bottom right. A black arrow points from the word "model" in the text above to the "PREDICTION" section of the journal cover.

SPECIAL SECTION PREDICTION

ESSAY

Data-driven predictions in the science of science

Aaron Clauset,^{1,2*} Daniel B. Larremore,² Roberta Sinatra^{3,4}

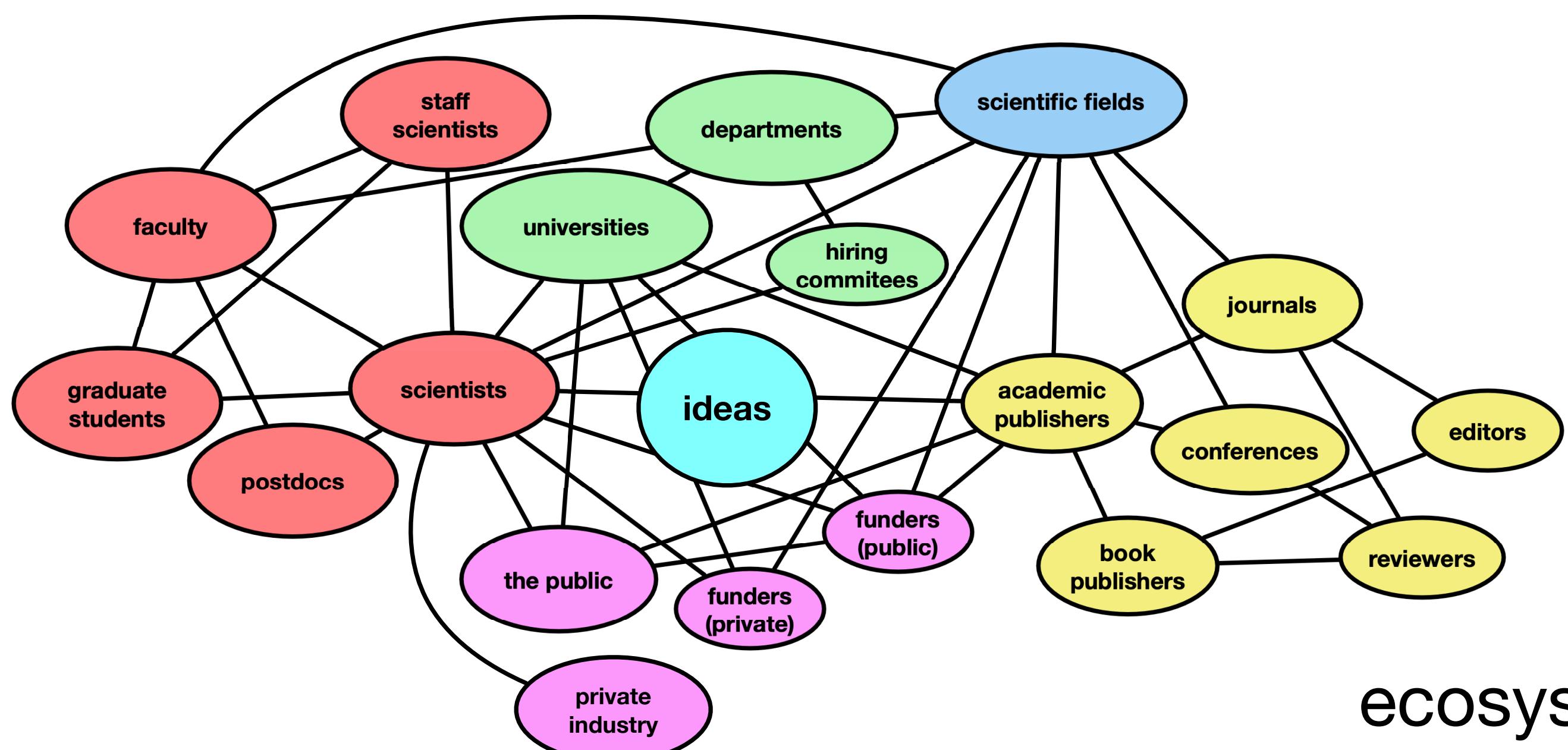
Science 355, 477-480 (2017)



science is an ecosystem*



diverse & messy ➔ scientists, journals / conferences, fields, universities, funders, etc
each playing different *ecological roles*, with interacting *priorities* and *incentives*

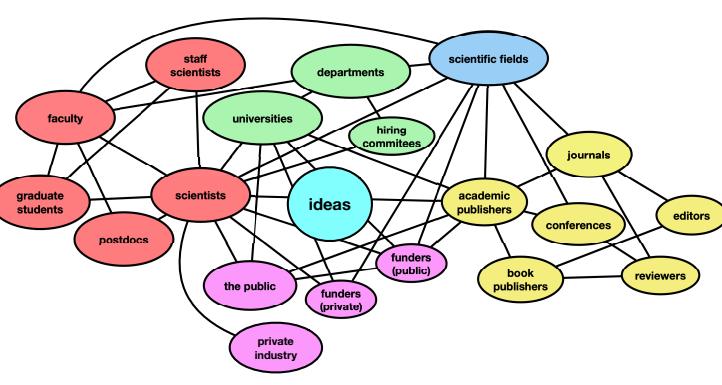


=
=



ecosystem dynamics

* this metaphor brings many useful ecological concepts : niches, communities, habitats, reproduction, extinction, speciation, famines, generalists vs. specialists, etc. and these concepts give us new and useful ways of thinking about the dynamics of the scientific community, bias, and the evolution of knowledge



science is an ecosystem



diverse & messy → scientists, journals / conferences, fields, universities, funders, etc

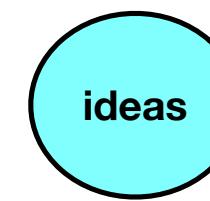
each playing different *ecological roles*, with interacting *priorities* and *incentives*

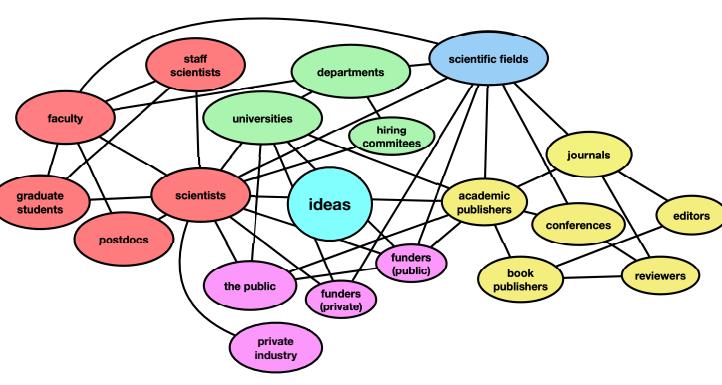
social biases somewhere in the ecosystem → complicated downstream effects
 [social disparities & epistemic inefficiencies]



gendered retention of faculty

- changes kind of grant proposals submitted
- changes kind of grants awarded
- changes kind of papers being written
- changes the kind of ideas studied





science is an ecosystem



diverse & messy → scientists, journals / conferences, fields, universities, funders, etc

each playing different *ecological roles*, with interacting *priorities* and *incentives*

social biases somewhere in the ecosystem → complicated downstream effects
 [social disparities & epistemic inefficiencies]

gendered retention of faculty

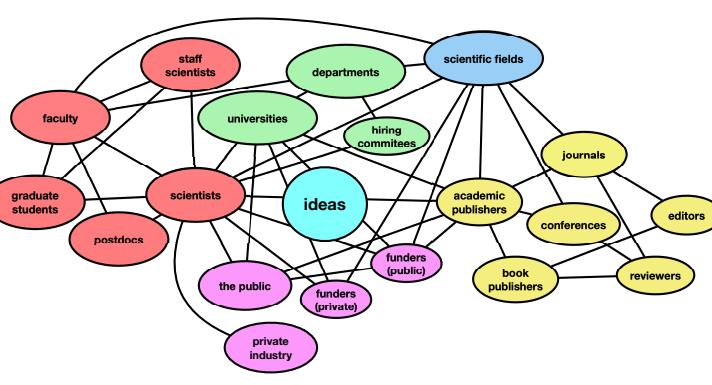
- changes kind of grant proposals submitted
- changes kind of grants awarded
- changes kind of papers being written
- changes the kind of ideas studied



biases in peer review

- changes kind of papers being published
- changes kind of grant proposals submitted
- changes kind of grants awarded
- changes who persists as scientists





science is an ecosystem



diverse & messy → scientists, journals / conferences, fields, universities, funders, etc

each playing different *ecological roles*, with interacting *priorities* and *incentives*

social biases somewhere in the ecosystem → complicated downstream effects
 [social disparities & epistemic inefficiencies]

gendered retention of faculty

- changes kind of grant proposals submitted
- changes kind of grants awarded
- changes kind of papers being written
- changes the kind of ideas studied

biases in peer review

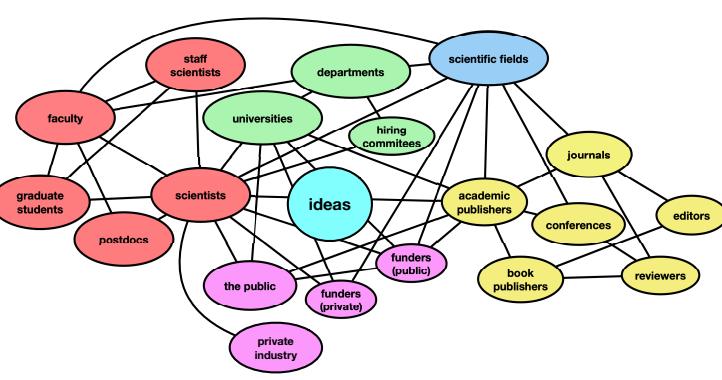
- changes kind of papers being published
- changes kind of grant proposals submitted
- changes kind of grants awarded
- changes who persists as scientists

hiring
committees

prestige bias in hiring

- changes who gets hired as faculty
- changes what ideas get worked on
- changes what ideas get funded
- changes what inventions get made

private
industry



science is an ecosystem



diverse & messy → scientists, journals / conferences, fields, universities, funders, etc

each playing different *ecological roles*, with interacting *priorities* and *incentives*

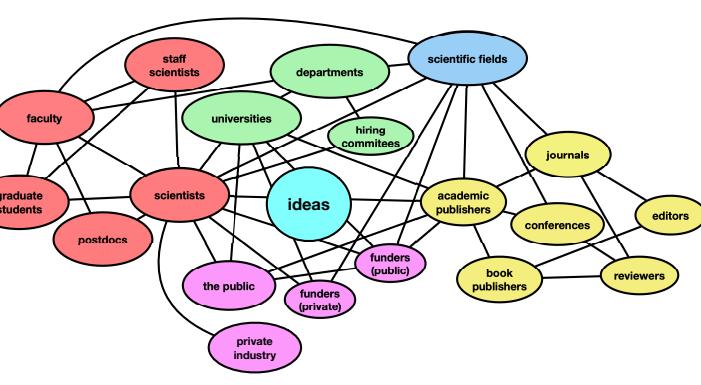
social biases somewhere in the ecosystem → complicated downstream effects

▶ difficult to improve one area independent of interactions with other areas

- over 1998-2003, US NIH budget \$14 billion → \$27 billion (2x)

- ecosystem effects : more time spent writing grant proposals, less time for research, success rate declined, average age of first R01 increased, pathway to biomedical tenure-track faculty jobs narrowed, etc.

▶ incentives and bias in one place can cause disparities downstream * ☹



science is an ecosystem



diverse & messy → scientists, journals / conferences, fields, universities, funders, etc
each playing different *ecological roles*, with interacting *priorities* and *incentives*
social biases somewhere in the ecosystem → complicated downstream effects
▶ difficult to improve one area independent of interactions with other areas

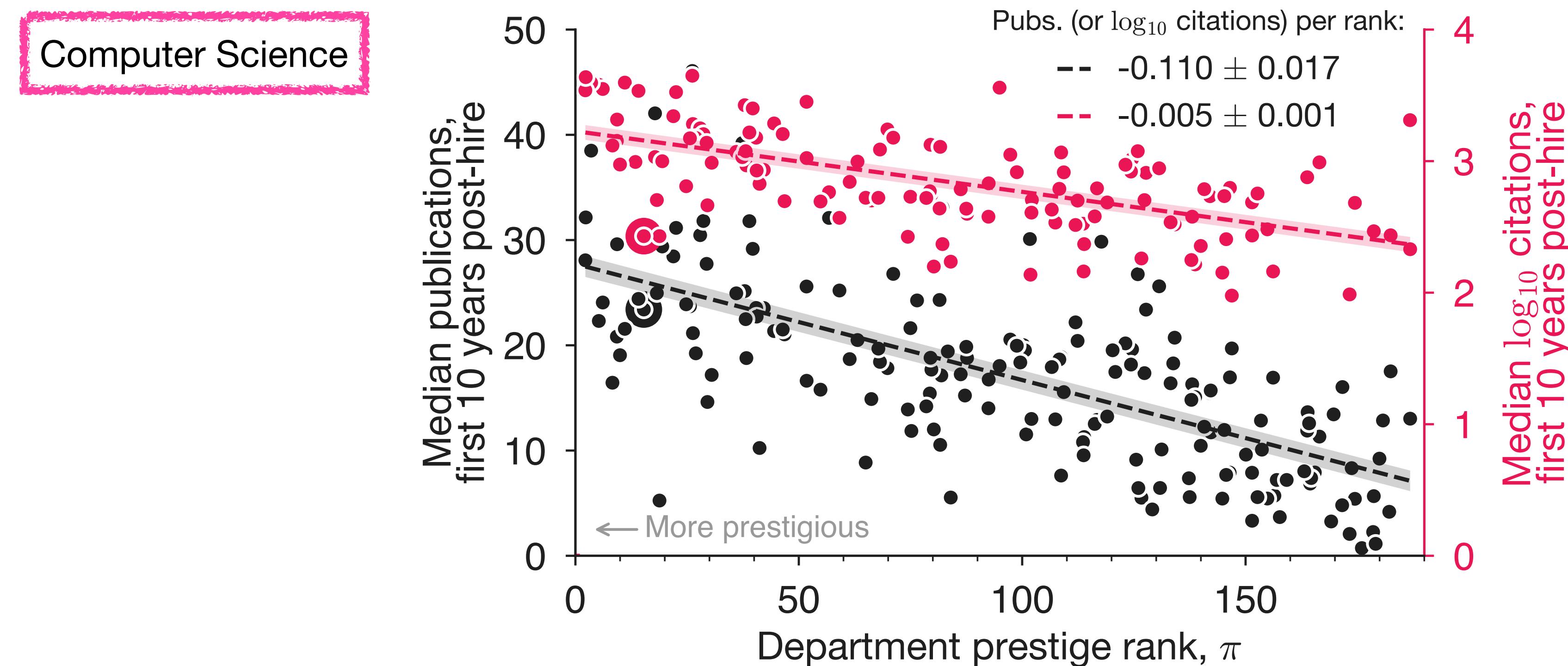
untangling this web of interactions requires
thinking about structural and network effects

→ in this talk, the productivity and
prominence of scientists

what drives productivity and prominence?



- ▶ scientific discourse is dominated by scientists trained + working at elite programs



what drives productivity and prominence?

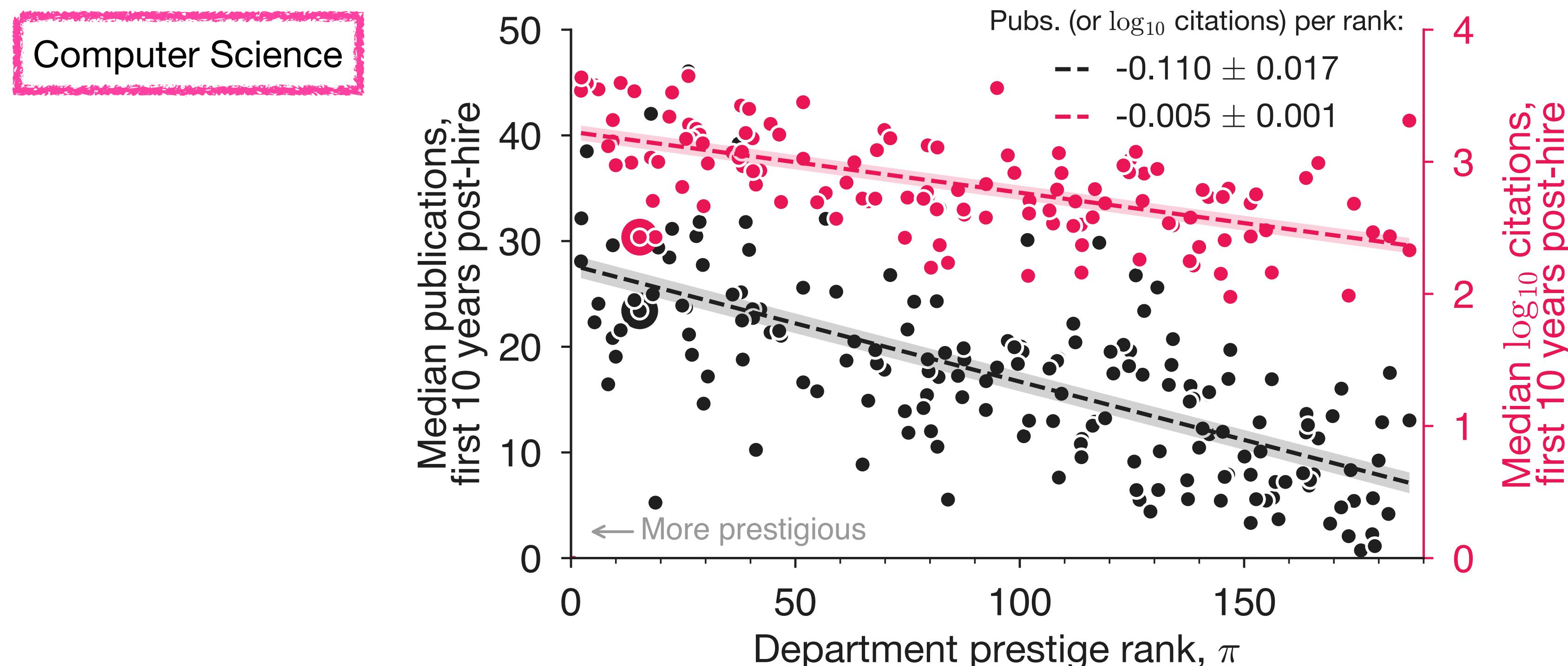
- ▶ scientific discourse is dominated by scientists trained + working at elite programs

but

elite training

elite faculty job

endogenous cumulative advantage → past achievements correlate with future achievements



okay if purely meritocratic.
but let's dig deeper:
how does this happen?

what drives productivity and prominence?

▶ scientific discourse is dominated by scientists trained + working at elite programs

two possible explanations of this correlation

idea 1: where a scientist trained.

- skill, talent, training, temperament, etc.

- elite programs select for natural potential for greatness (apparent merit)

- e.g., most productive scientists have elite pedigree, Harvard, Yale, Oxford, Stanford, etc.

where you **train** causes future productivity and prominence

what drives productivity and prominence?

▶ scientific discourse is dominated by scientists trained + working at elite programs

two possible explanations of this correlation

idea 2: where a scientist works.

- environmental factors, resources, people, support

- beyond a basic training, a scientist's output is driven by local environment

- e.g., moving from poor to rich environments improves output, and vice versa

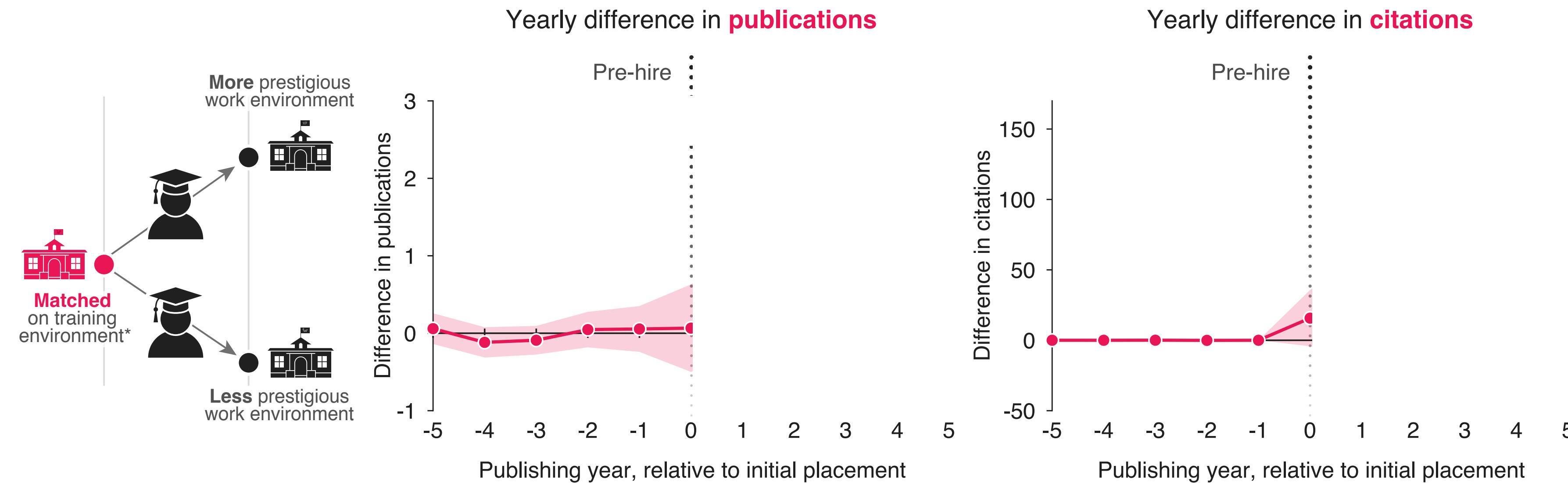
where you work causes future productivity and prominence

what drives productivity and prominence?

a test: faculty hiring is a quasi-natural experiment

2,453 early career CS faculty hires, and their 200,000 publications & 7.4M citations

match faculty pairs on *similar training, different placement*



caliper matched faculty on {gender, subfield, hiring prestige OR phd prestige, year of placement, postdoctoral training}. results robust to caliper variations

publications: N = 196 pairs, p < 0.005 (t-test); citations: N = 96 pairs, p < 0.005 (t-test)

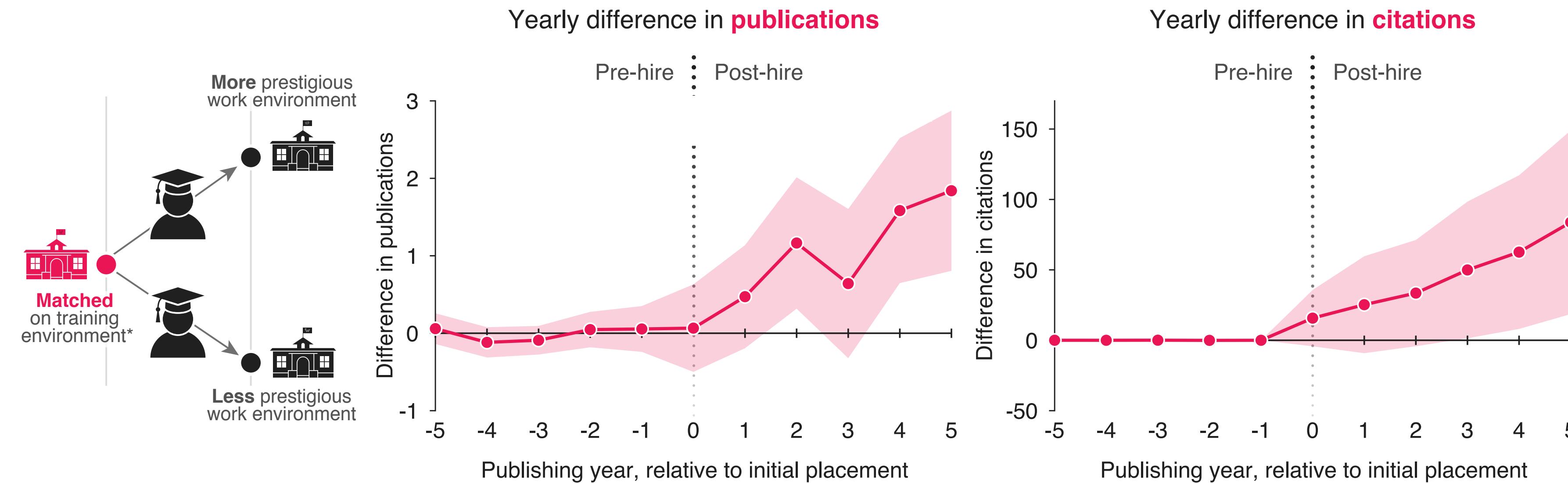
Way et al. ["Productivity, prominence, and the effects of academic environment"](#) (2019)

what drives productivity and prominence?

a test: faculty hiring is a quasi-natural experiment

2,453 early career CS faculty hires, and their 200,000 publications & 7.4M citations

match faculty pairs on similar training, different placement → *different productivity & prominence*



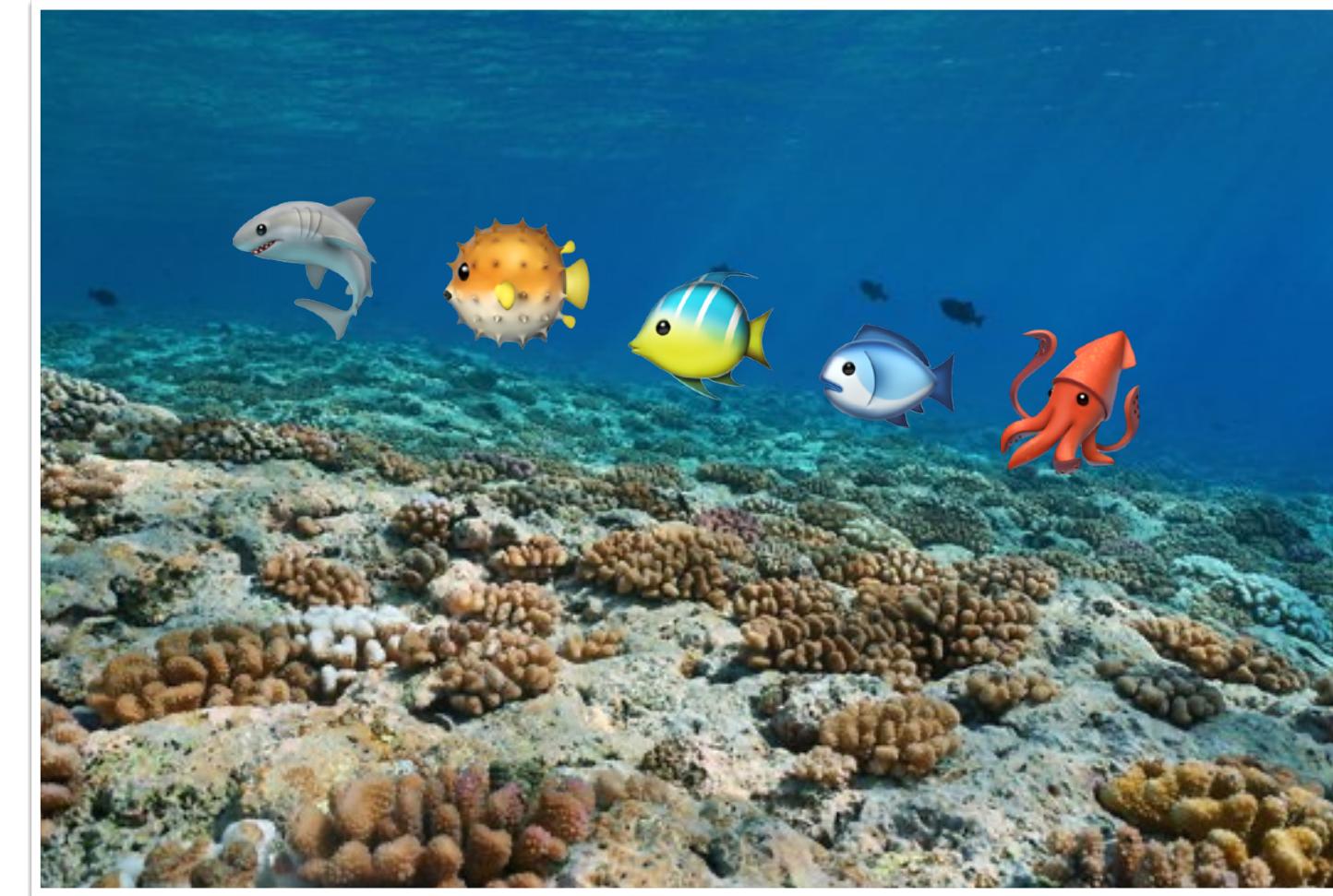
caliper matched faculty on {gender, subfield, hiring prestige OR phd prestige, year of placement, postdoctoral training}. results robust to caliper variations

publications: N = 196 pairs, p < 0.005 (t-test); citations: N = 96 pairs, p < 0.005 (t-test)

Way et al. "Productivity, prominence, and the effects of academic environment" (2019)

what drives productivity and prominence?

- ▶ prestige is a *structural variable* in the scientific ecosystem → like a sloping seafloor in a coral reef
 - elite training does not drive scholarly productivity or prominence*
 - working environment—not training—appears to drive your output



prestige rules everything around me

what drives productivity and prominence?

- ▶ prestige is a *structural variable* in the scientific ecosystem → like a sloping seafloor in a coral reef
 - elite training does not drive scholarly productivity or prominence*
 - working environment—not training—appears to drive your output

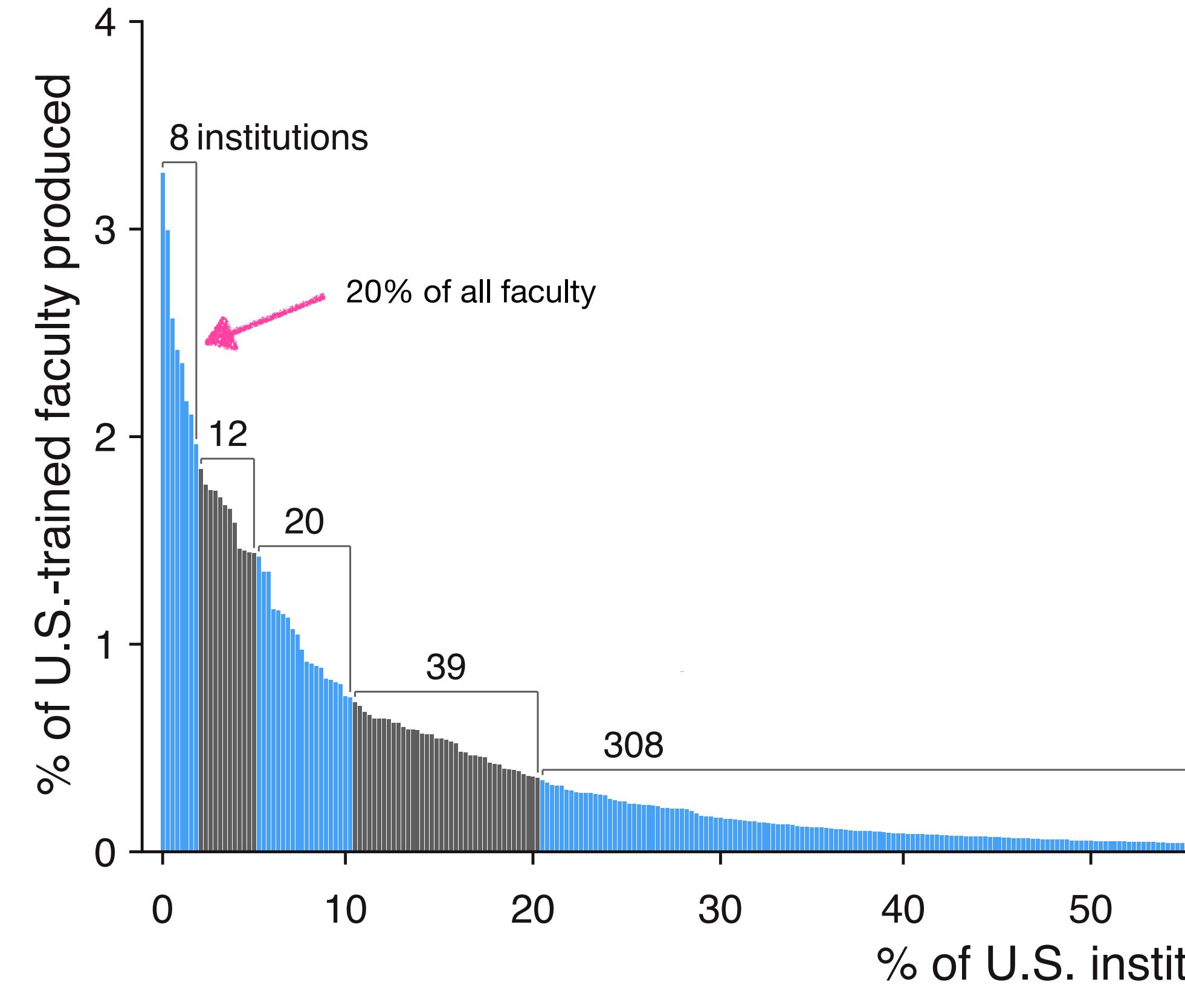
■ why do elite institutions dominate science?

doctoral prestige → faculty location

Clauset et al. (2015), Wapman et al. (2022)

80% of faculty trained at 20% of universities

elite programs shape everyone's research agendas



what drives productivity and prominence?

- ▶ prestige is a *structural variable* in the scientific ecosystem → like a sloping seafloor in a coral reef
 - elite training does not drive scholarly productivity or prominence*

working environment—not training—appears to drive your output

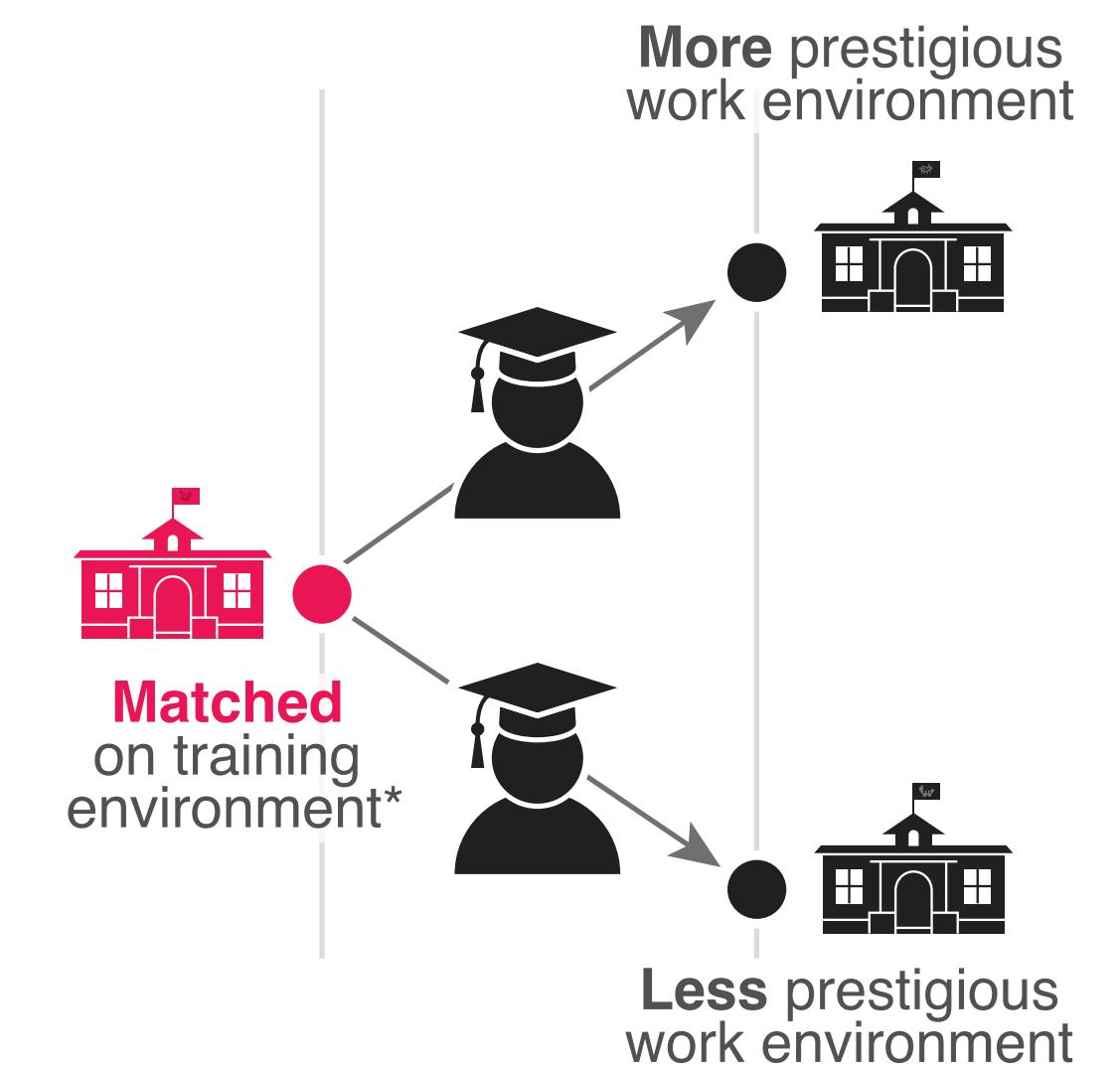
- why do elite institutions dominate science?

doctoral prestige → faculty location → scholarly output

these results, Way et al. (2019)

- working environment causes higher output
- an *environmental* mechanism for the cumulative advantage of elite scientists

but how?

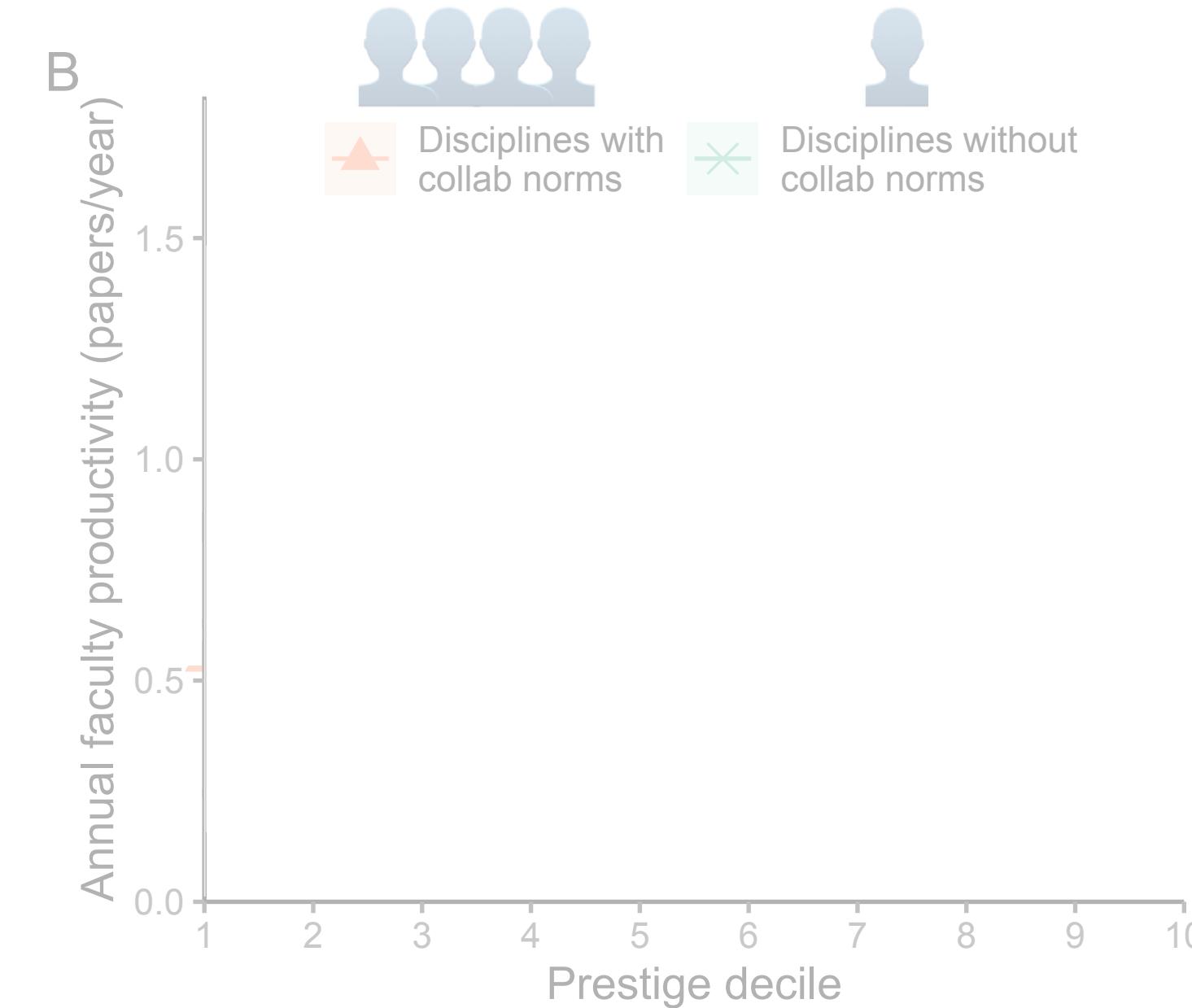
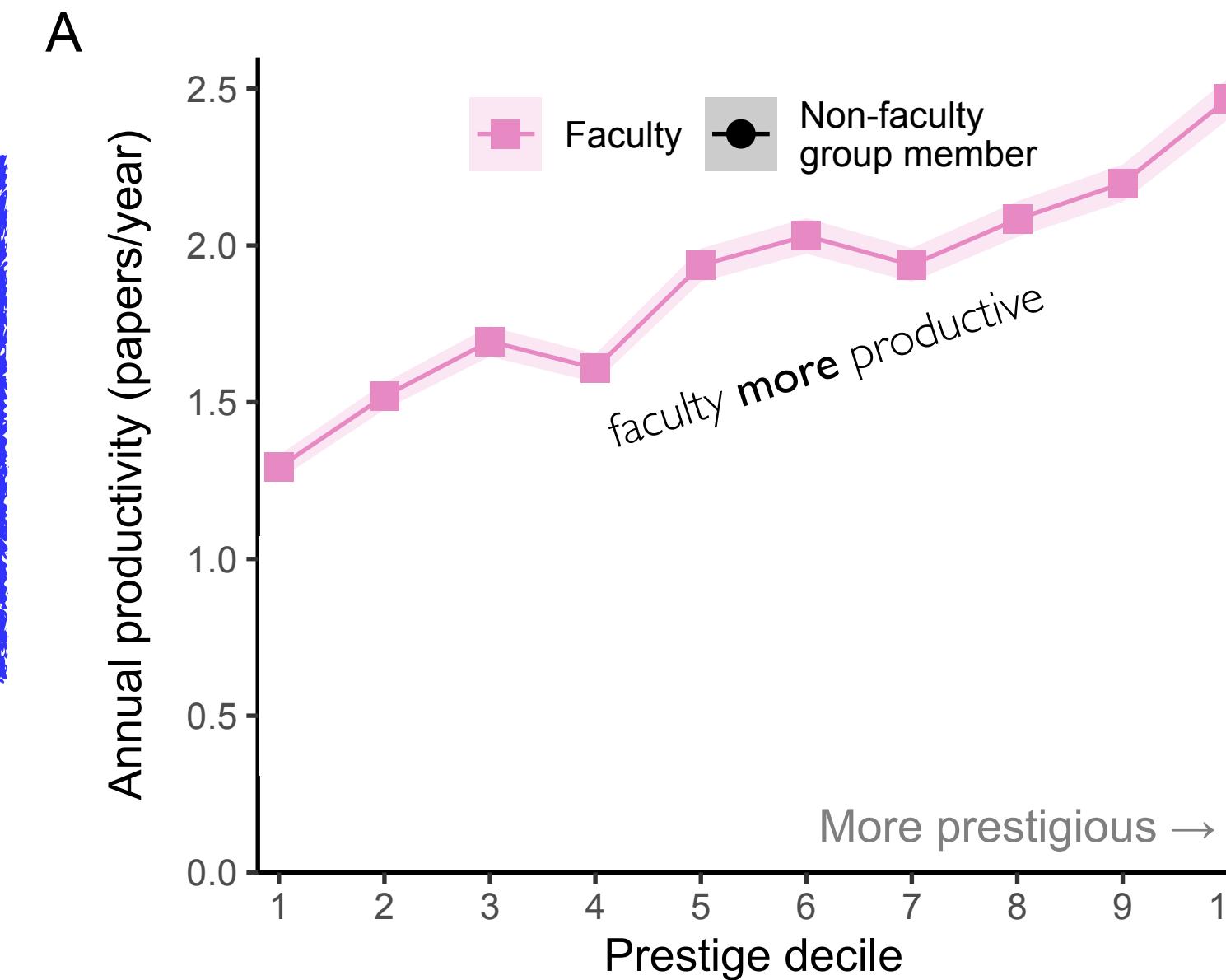


a labor advantage

prestige → larger research groups → productivity

- ▶ data on 78,802 regular faculty across 26 fields at 262 U.S. PhD-granting institutions
+ complete publication data (Web of Science) + divide fields by coauthorship norms

faculty productivity

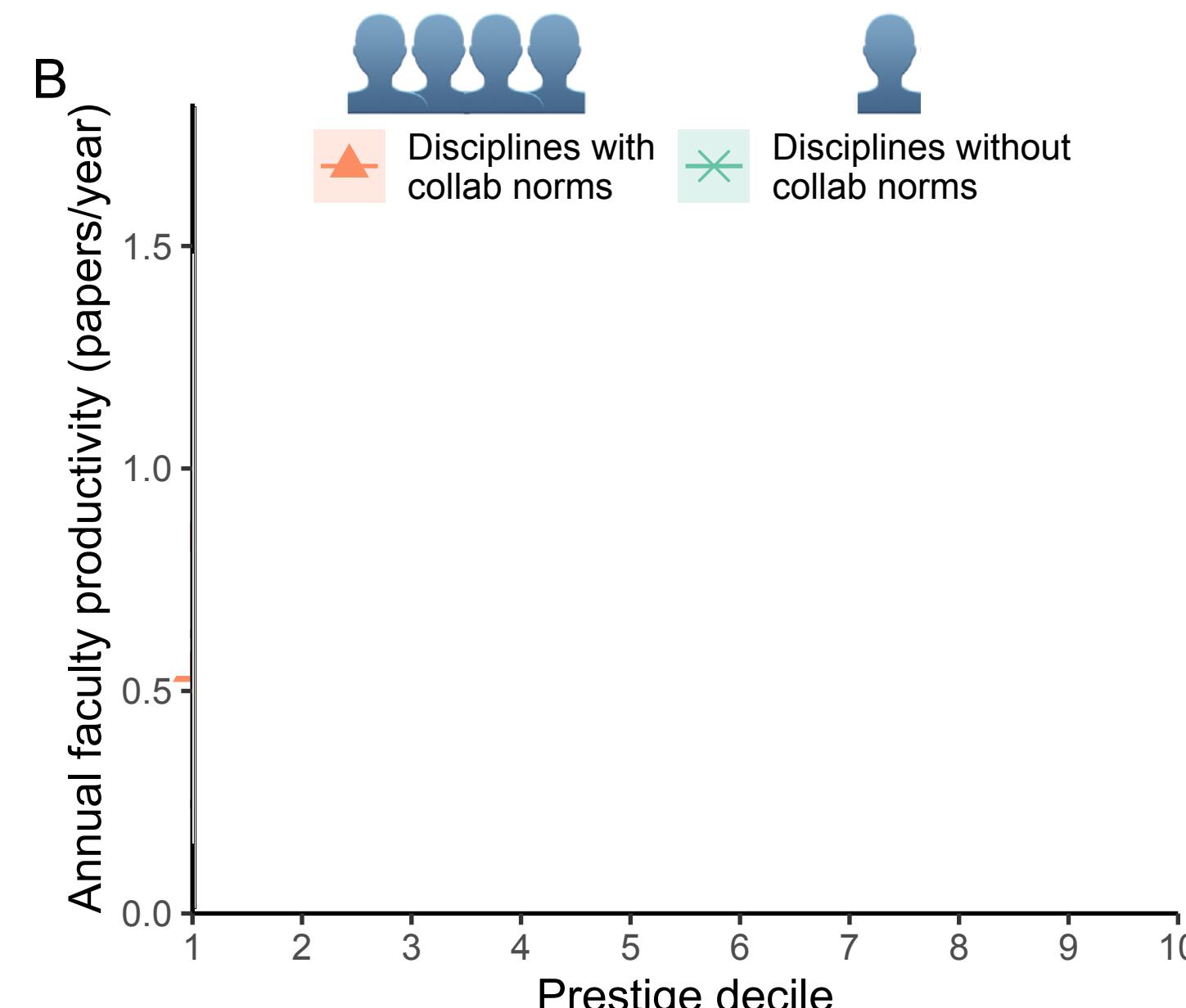
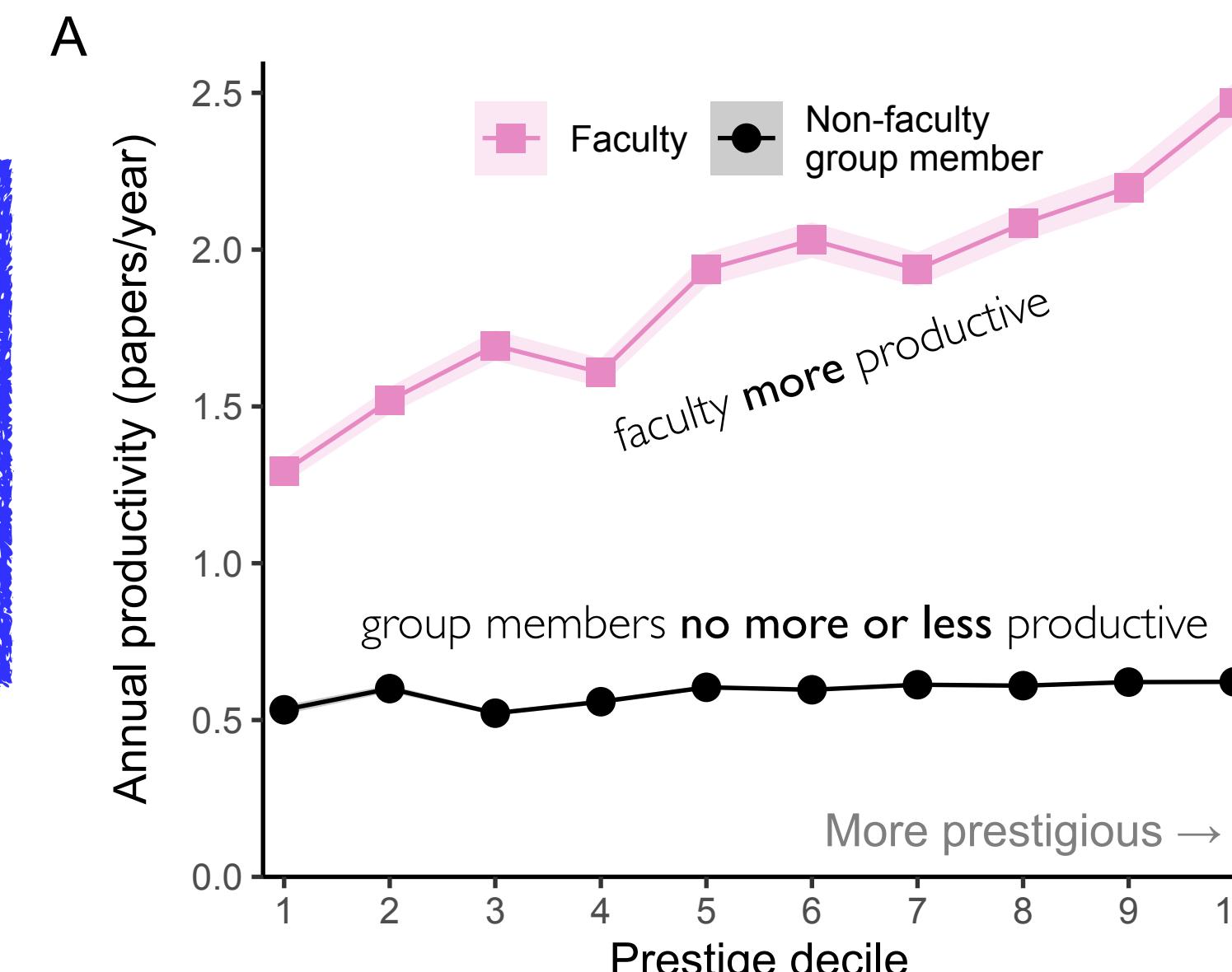


a labor advantage

prestige → larger research groups → productivity

- ▶ data on 78,802 regular faculty across 26 fields at 262 U.S. PhD-granting institutions + complete publication data (Web of Science) + divide fields by coauthorship norms

faculty productivity =
group + individual



fields without
coauthorship norms
should see weak/no
prestige-productivity
effect

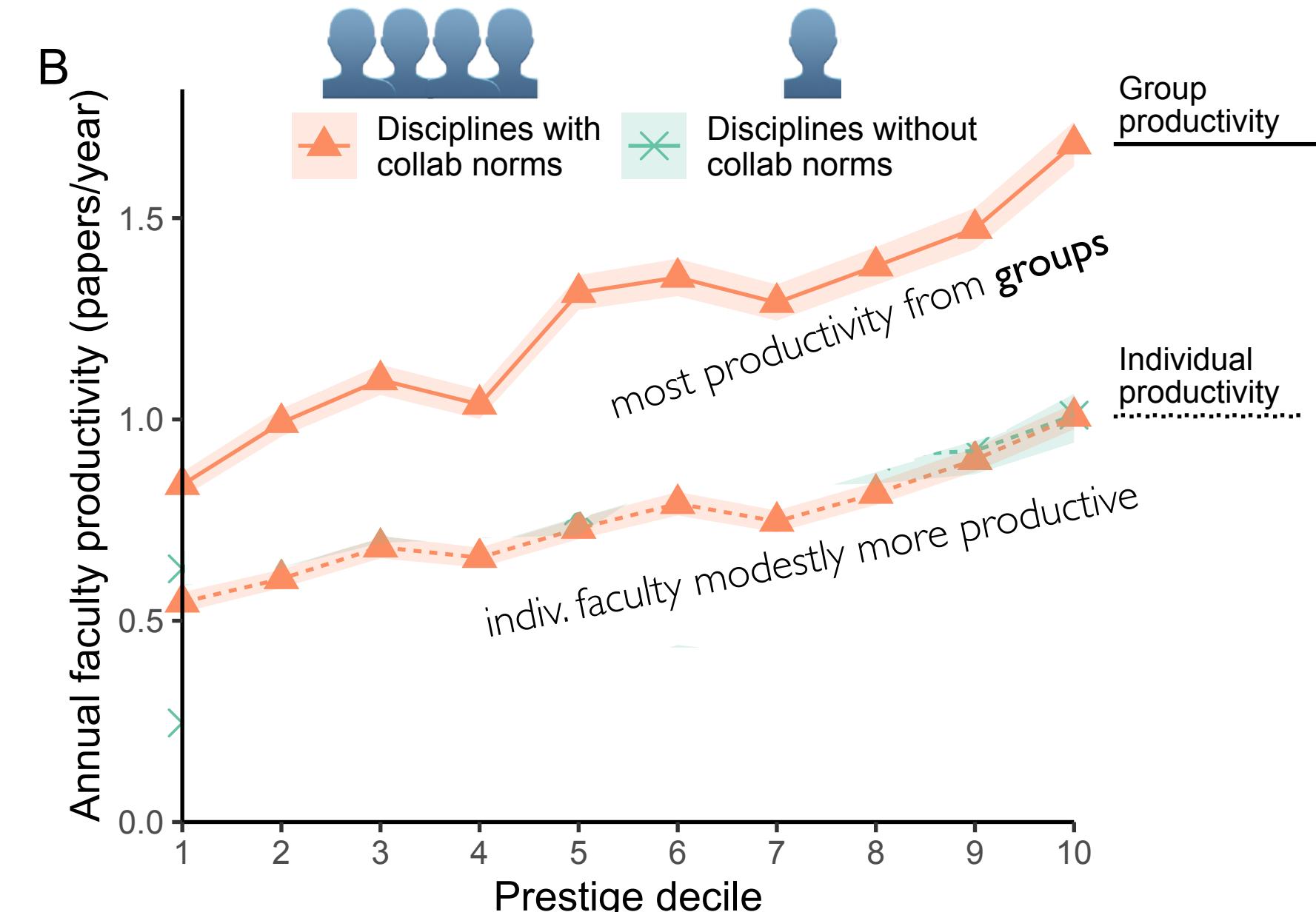
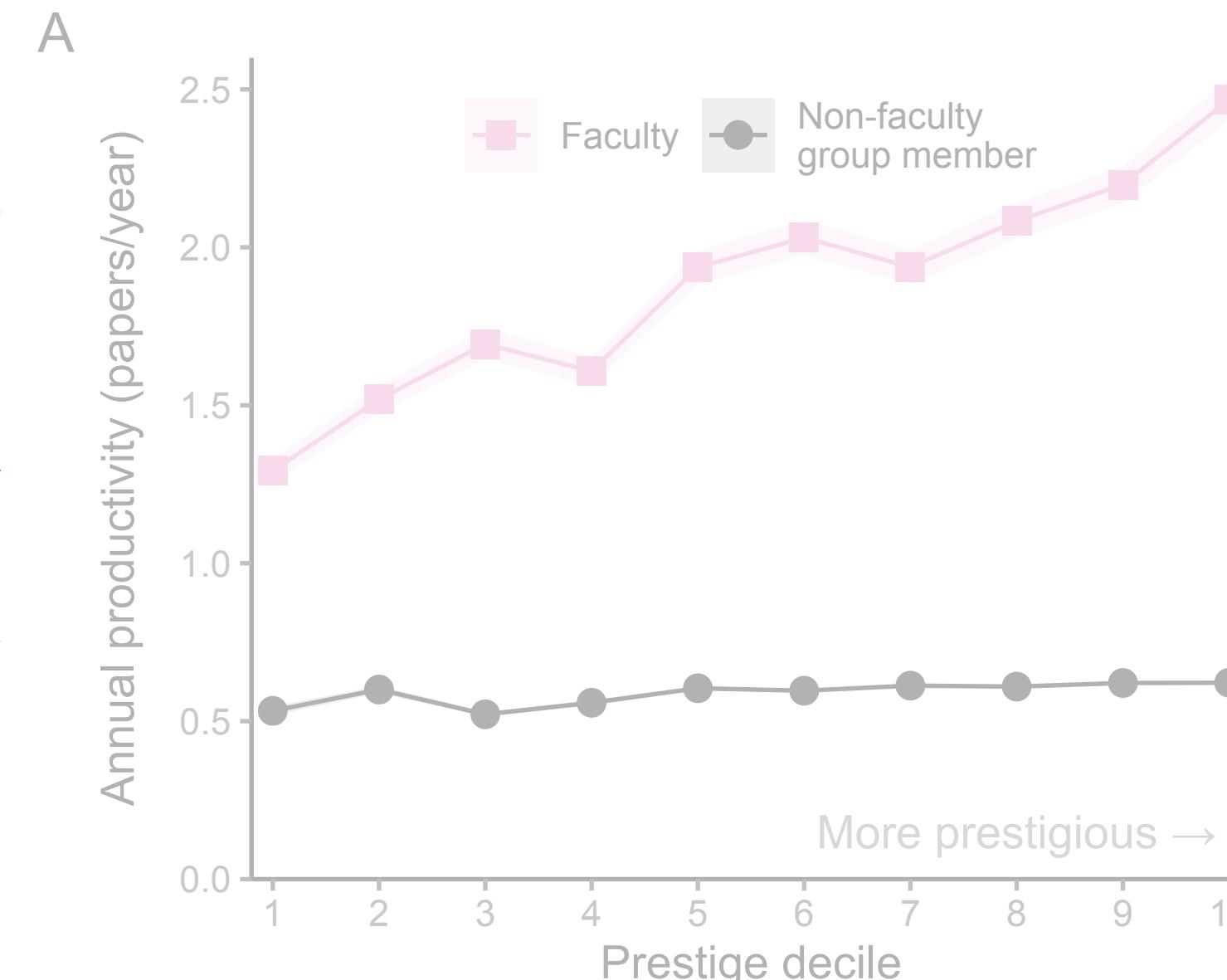
a labor advantage

prestige → larger research groups → productivity

data on 78,802 regular faculty across 26 fields at 262 U.S. PhD-granting institutions
+ complete publication data (Web of Science) + divide fields by coauthorship norms

▶ fields with coauthorship norms = strong prestige-productivity effect

faculty productivity =
group + individual



group vs. individual productivity estimated from bibliographic records: does pub have a same-institution early-career coauthor?
there's simply *much more* available labor at elite institutions (data from US National Science Foundation)
Zhang et al., "Labor advantages drive the greater productivity of faculty at elite universities" (2022)

Disciplines with collaboration norms	Disciplines without collaboration norms
Biological Sciences	Economics
Engineering	Mathematical Sciences
Medical Sciences	Language, Literature, Culture
Psychological Sciences	Political Science
Physical Sciences	History
Chemical Sciences	Anthropology
Computational Science	Sociology
Health	Education
Business	Philosophy
Earth Sciences	Arts
Agriculture	Theology and Religion
Architecture, Design, Planning	Geography
	Linguistics

a labor advantage

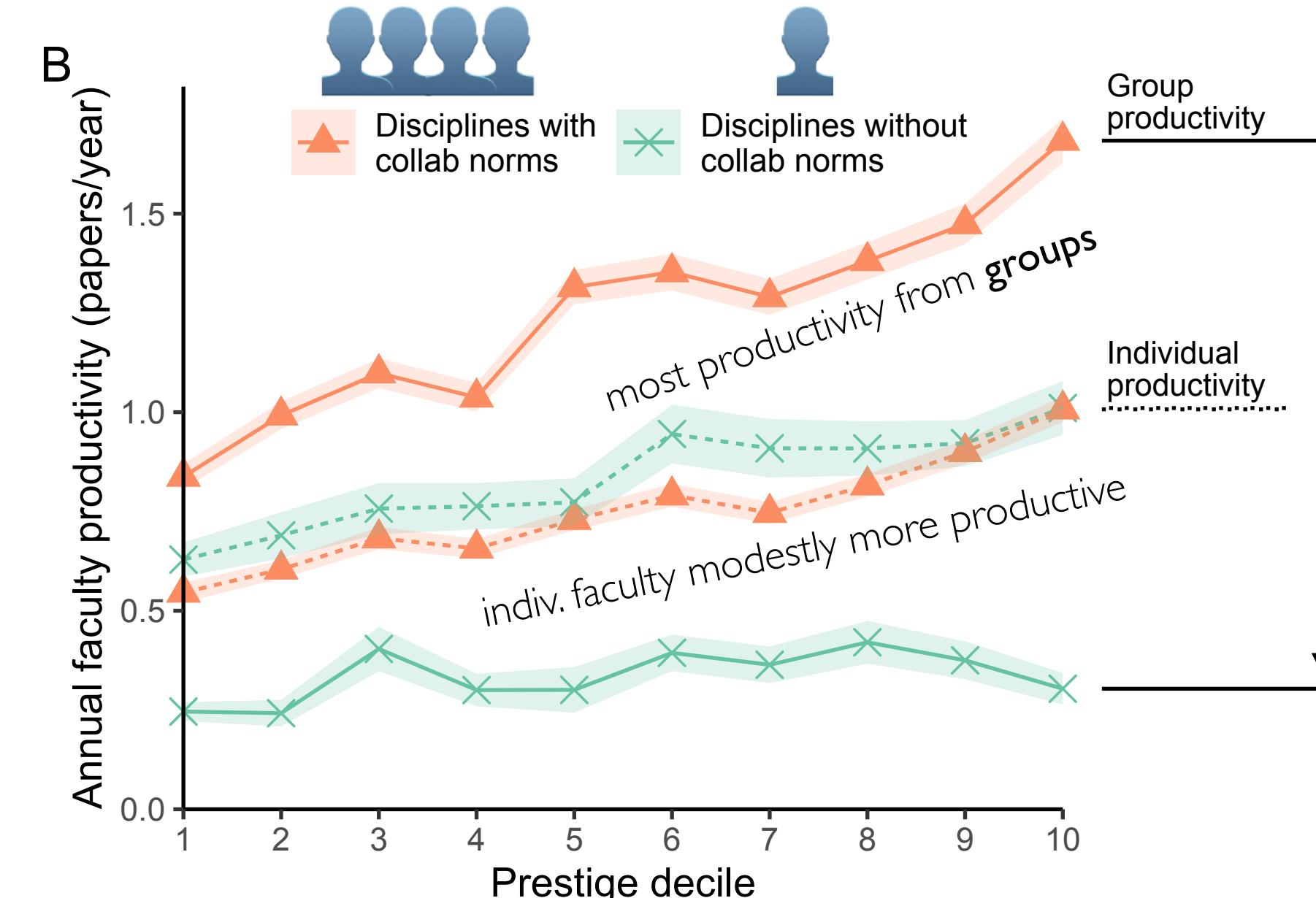
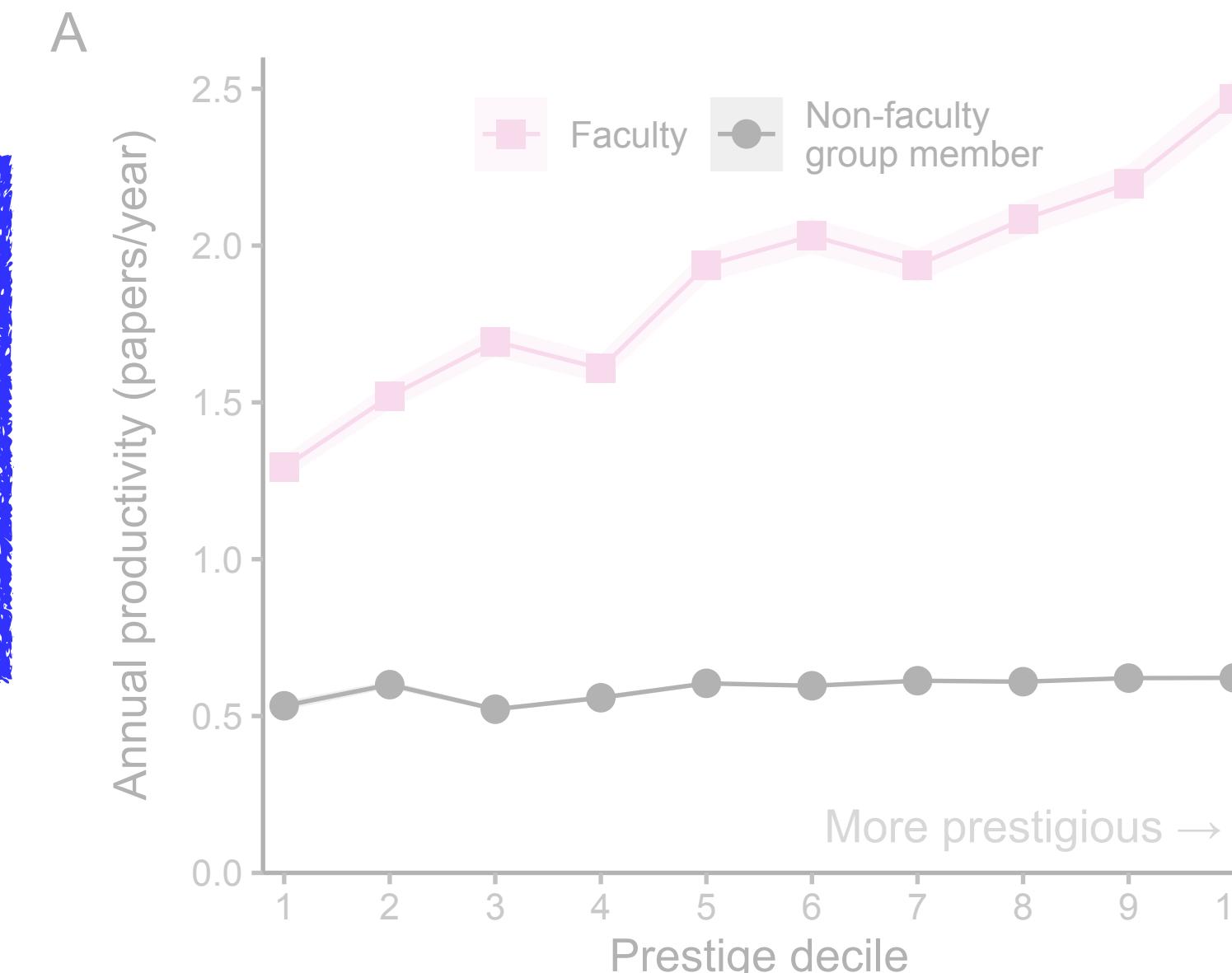
prestige → larger research groups → productivity

data on 78,802 regular faculty across 26 fields at 262 U.S. PhD-granting institutions
+ complete publication data (Web of Science) + divide fields by coauthorship norms

fields with coauthorship norms = strong prestige-productivity effect

► fields without coauthorship norms = much weaker effect

faculty productivity =
group + individual



group vs. individual productivity estimated from bibliographic records: does pub have a same-institution early-career coauthor?

there's simply *much more* available labor at elite institutions (data from US National Science Foundation)

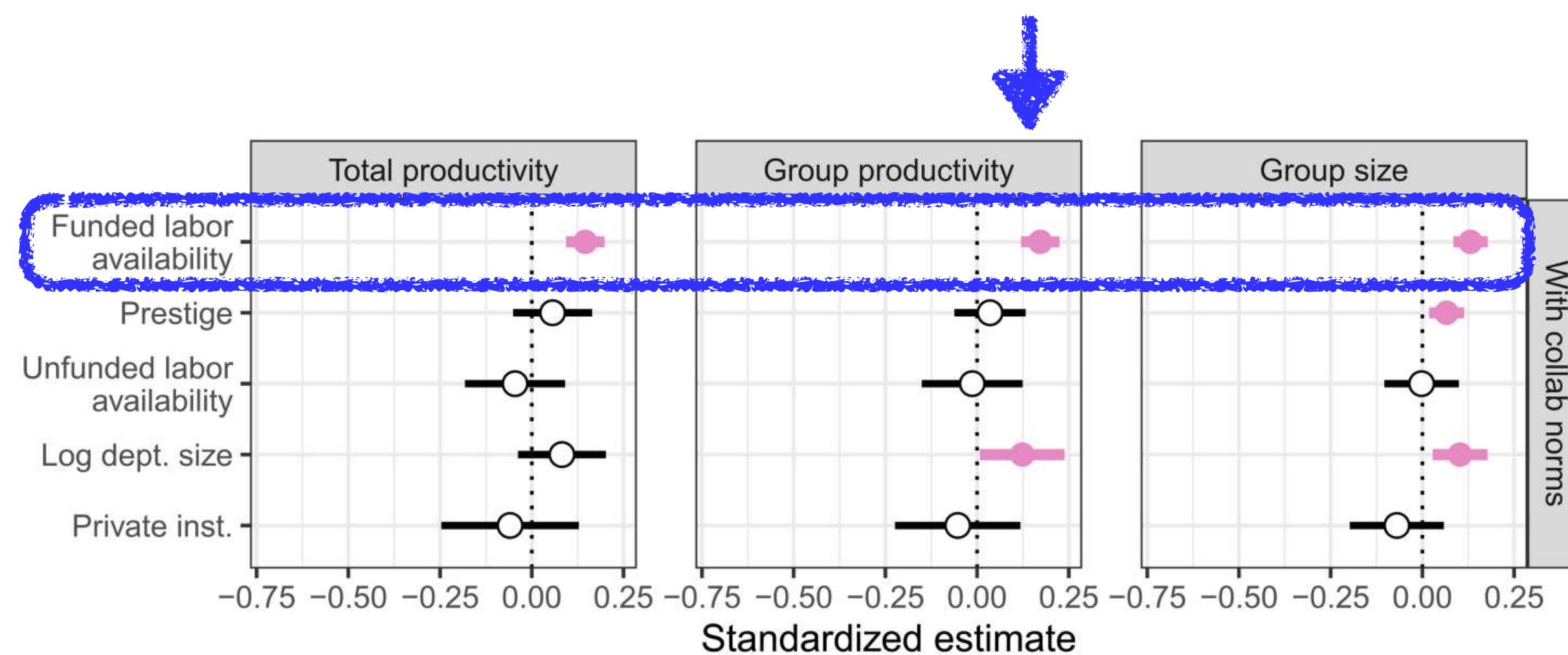
Zhang et al., "Labor advantages drive the greater productivity of faculty at elite universities" (2022)

Disciplines with collaboration norms	Disciplines without collaboration norms
Biological Sciences	Economics
Engineering	Mathematical Sciences
Medical Sciences	Language, Literature, Culture
Psychological Sciences	Political Science
Physical Sciences	History
Chemical Sciences	Anthropology
Computational Science	Sociology
Health	Education
Business	Philosophy
Earth Sciences	Arts
Agriculture	Theology and Religion
Architecture, Design, Planning	Geography
	Linguistics

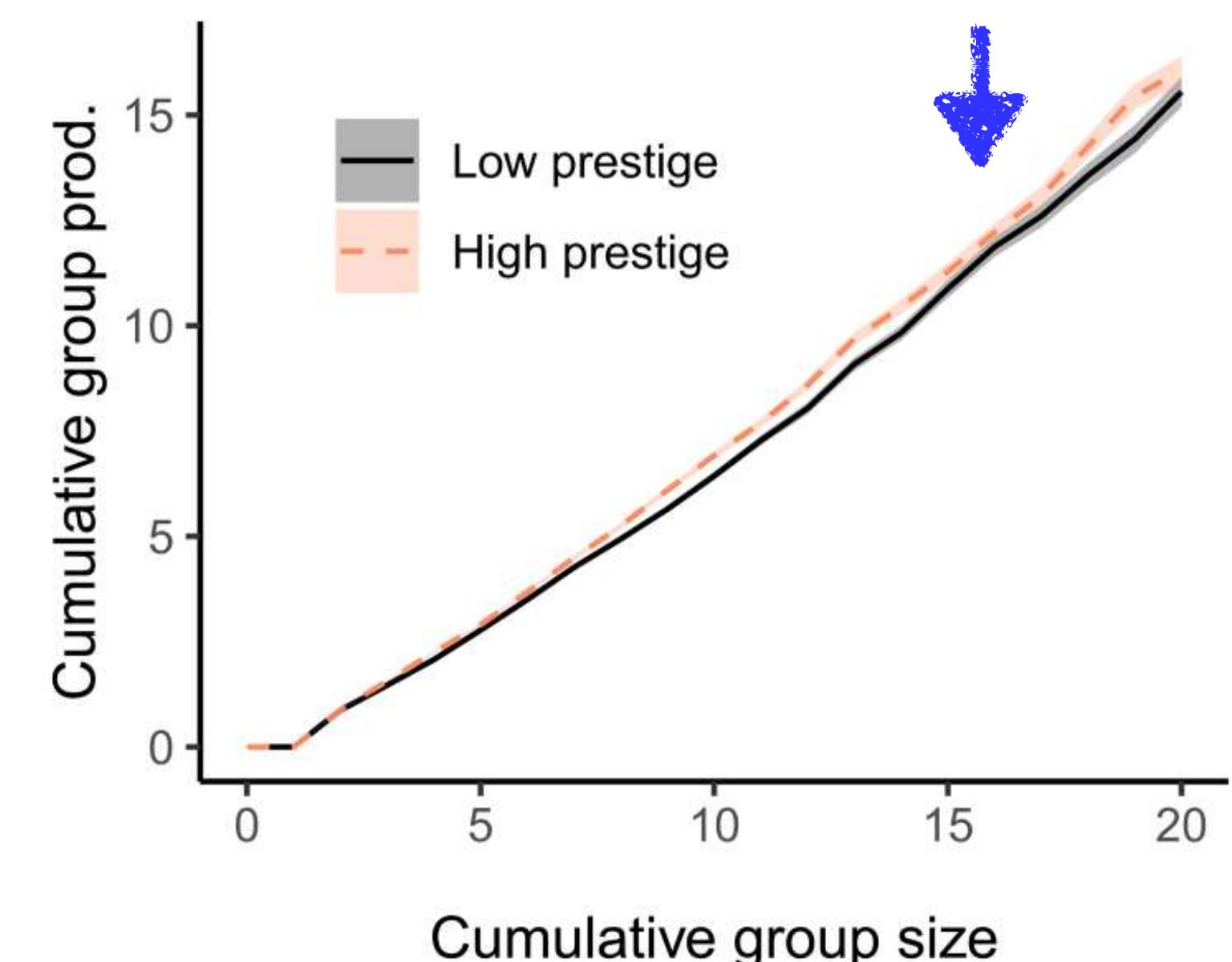
a labor advantage

prestige → larger research groups → productivity

available labor is the main variable for predicting productivity & group size



productivity grows at same rate
with group size at high & low
prestige institutions

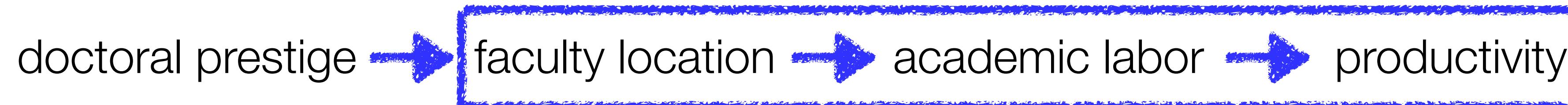


what drives productivity and prominence?

▶ prestige is a *structural variable* in the scientific ecosystem

working environment (not pedigree) drives productivity, primarily via a *labor advantage*

■ why do elite institutions dominate science?



■ more labor = more coauthors

in a network, more coauthors = higher node degree
how much does it matter who you collaborate with?

