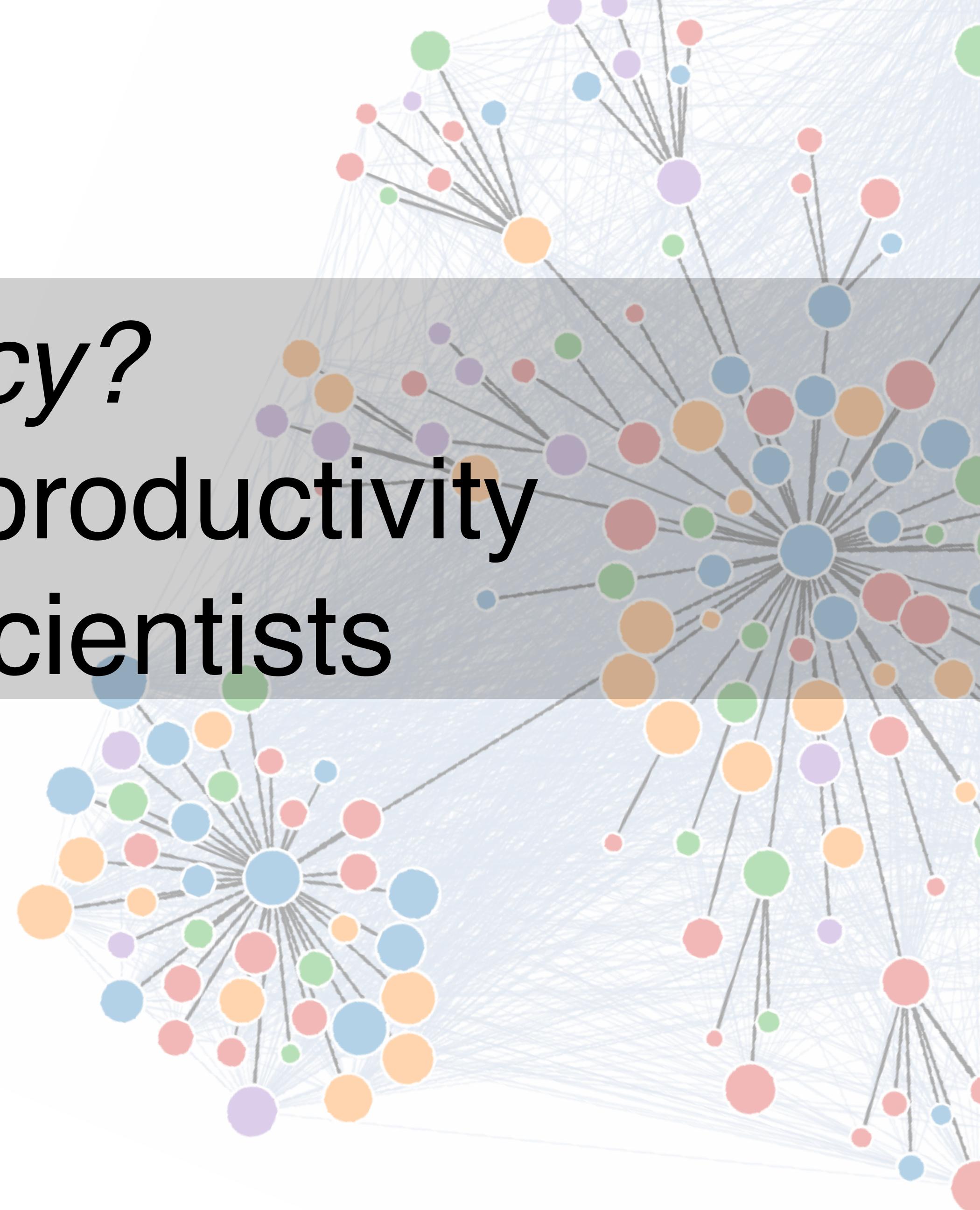


How much of a meritocracy?

Untangling the drivers of productivity and prominence among scientists

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



desire to predict discovery is pervasive

- ▶ **what will be discovered?**
- ▶ **by whom, when, and where?**



desire to predict discovery is pervasive

- ▶ what will be discovered?
- ▶ by whom, when, and where?



individuals	what questions are useful, impactful, fundable?
publishers, funders	what manuscripts or projects will be most impactful?
hiring committees	which applicant will perform best? which will make most valuable contributions?
society	how can tax and other dollars be invested to make technological, biomedical, and scientific advances?

toward a science of science

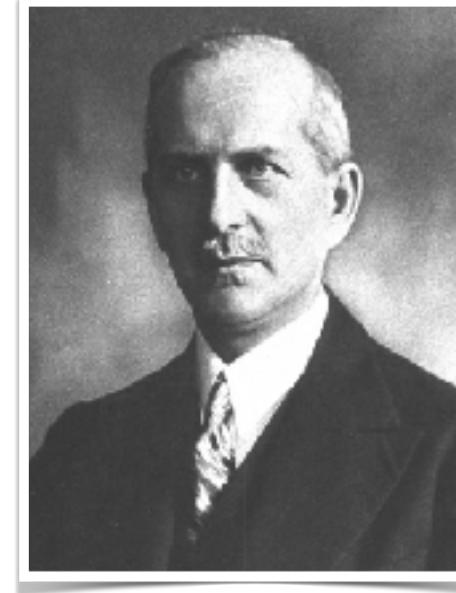
- ▶ **simple idea, with a 150+ year history**

toward a science of science

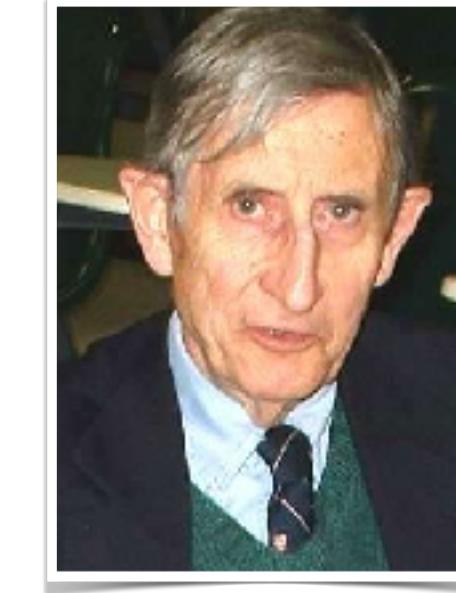
▶ simple idea, with a 150+ year history



Bolesław Prus
(1847-1912)



Florian Znaniecki
(1882-1958)



Freeman Dyson
(1923-2020)



Steven Weinberg
(1933-2021, Nobel Physics 1979)



Harriet Zuckerman
(1937-)

...
...

- philosophy, physics, sociology, informatics, ...
- mainly conceptual, focusing on goals and general approaches
(Weinberg: "to explain the world") (Dyson: "birds and frogs")
- progress toward a genuine "science of science" was slow
 - hard to get good data
 - judgment of experts seemed good enough

toward a science of science

the state of the field

toward a science of science

the state of the field

- ▶ *abundant but crude data*
publications + citations + people + funding
- Google Scholar, PubMed, Web of Science, arXiv, JSTOR, ORCID, EasyChair, NIH, NSF, patents, academic CVs, etc.

APS Data Sets for Research

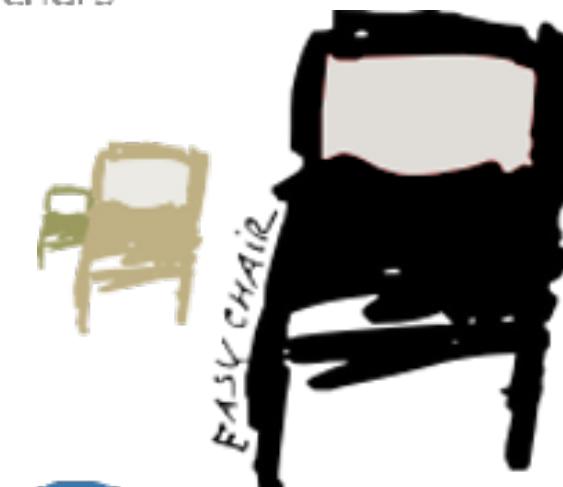
WEB OF SCIENCE™



ORCID

Connecting Research
and Researchers

arXiv.org



PubMed

toward a science of science

the state of the field

▶ *abundant but crude data*

publications + citations + people + funding

- Google Scholar, PubMed, Web of Science, arXiv, JSTOR, ORCID, EasyChair, NIH, NSF, patents, academic CVs, etc.

▶ *piecemeal theories*

- few predictive models, lots of descriptive work, complicated social dynamics, an ecosystem of actors

APS Data Sets for Research

WEB OF SCIENCE™



ORCID

Connecting Research
and Researchers

arXiv.org



PubMed

toward a science of science

the state of the field

▶ *abundant but crude data*

publications + citations + people + funding

- Google Scholar, PubMed, Web of Science, arXiv, JSTOR, ORCID, EasyChair, NIH, NSF, patents, academic CVs, etc.

▶ *piecemeal theories*

- few predictive models, lots of descriptive work, complicated social dynamics, an ecosystem of actors

▶ *growing interdisciplinary community*

- computer scientists, information scientists, economists, sociologists, statisticians, physicists, biologists, etc.

🎉 an exciting time to do science of science! 🎉

APS Data Sets for Research

WEB OF SCIENCE™



ORCID

Connecting Research
and Researchers

arXiv.org



PubMed

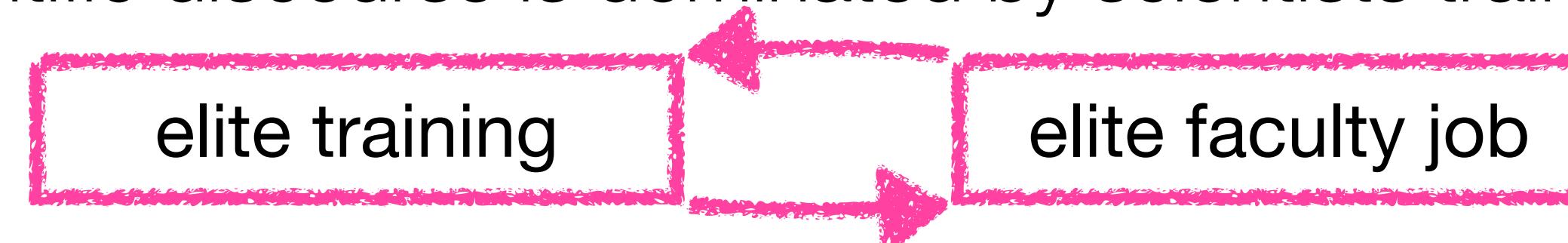
what drives productivity and prominence?

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

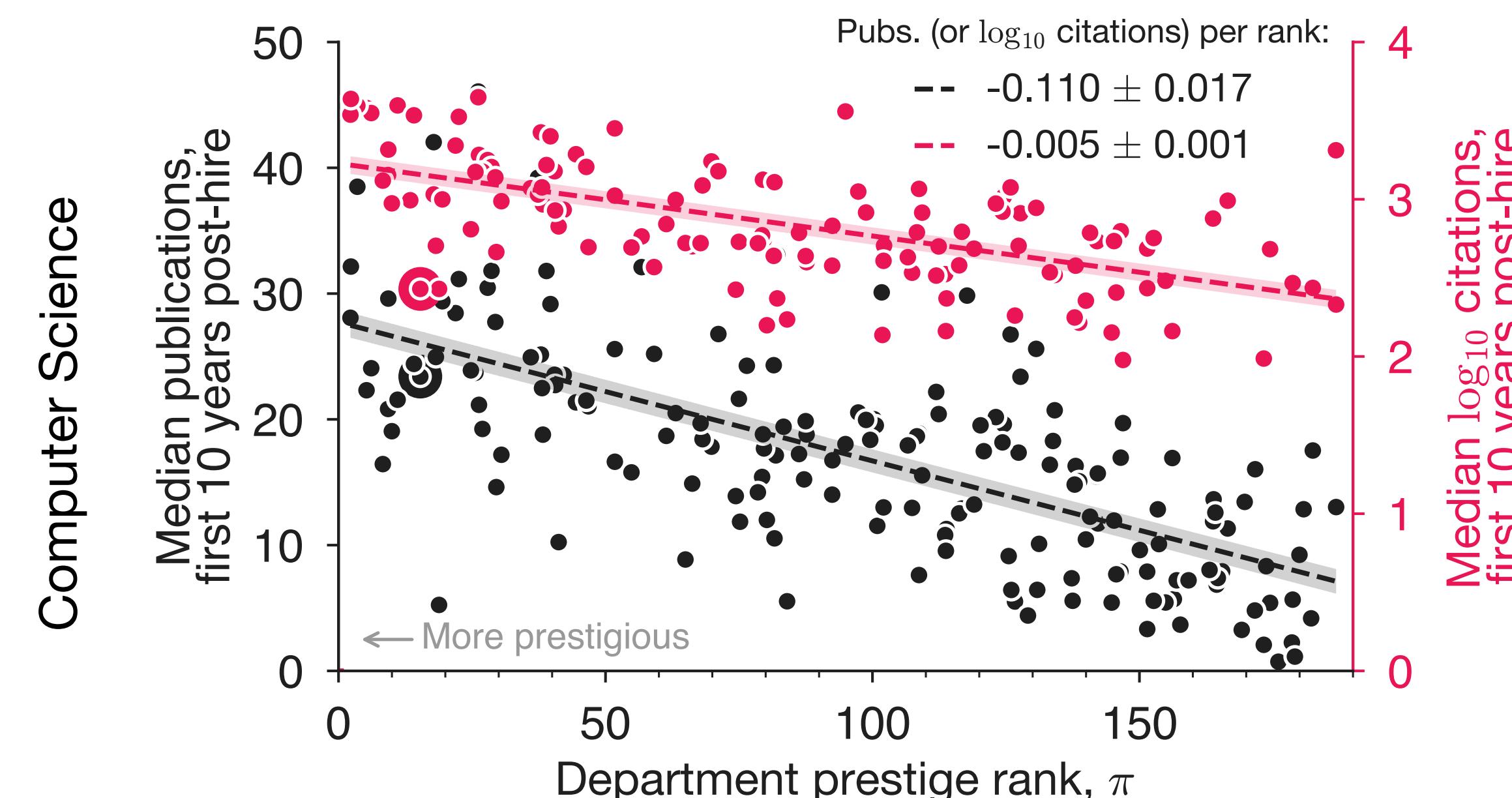
what drives productivity and prominence? 🧙

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

but



endogenous cumulative advantage → past achievements correlate with future achievements



what drives productivity and prominence?

- ▶ scientific discourse is dominated by scientists trained + working at elite programs
- two possibilities explain this correlation
- idea 1: where a scientist trained.

what drives productivity and prominence?



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

two possibilities explain this correlation

idea 1: where a scientist trained.

- skill, talent, training, temperament, etc.
- faculty hiring sorts people by their natural potential for good outcomes
- e.g., most productive scientists have elite pedigree, Harvard, Yale, Penn, 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 possibilities explain this correlation
- idea 2: where a scientist works.

what drives productivity and prominence?

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

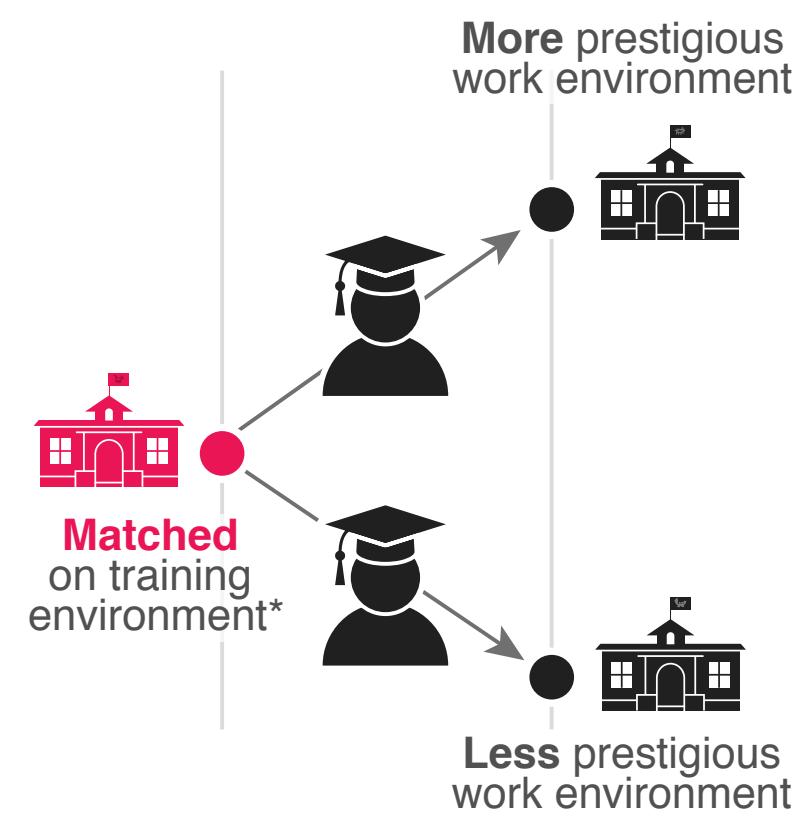


what drives productivity and prominence?

a test: faculty hiring is a quasi-natural experiment

early career CS faculty placements + DBLP matched productivity + Google Scholar citations

matching faculty on *similar training, different placement*

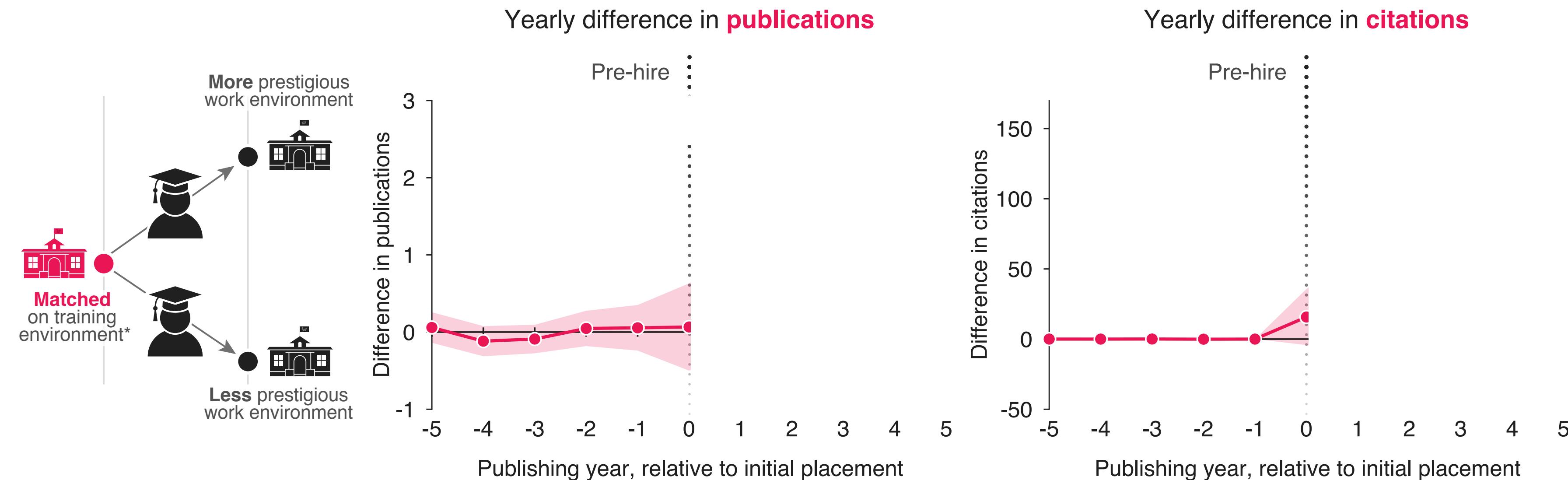


what drives productivity and prominence?

a test: faculty hiring is a quasi-natural experiment

early career CS faculty placements + DBLP matched productivity + Google Scholar citations

matching faculty 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)

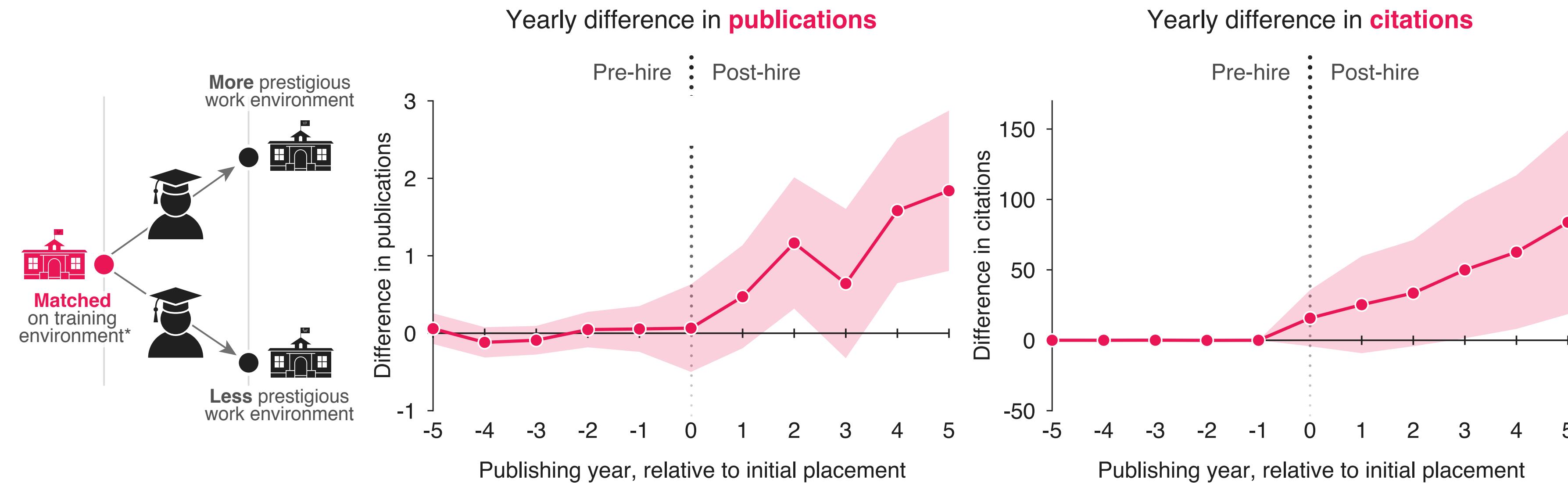
Way et al. "Productivity, prominence, and the effects of academic environment" (2019)

what drives productivity and prominence?

a test: faculty hiring is a quasi-natural experiment

early career CS faculty placements + DBLP matched productivity + Google Scholar citations

matching faculty 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 science of science

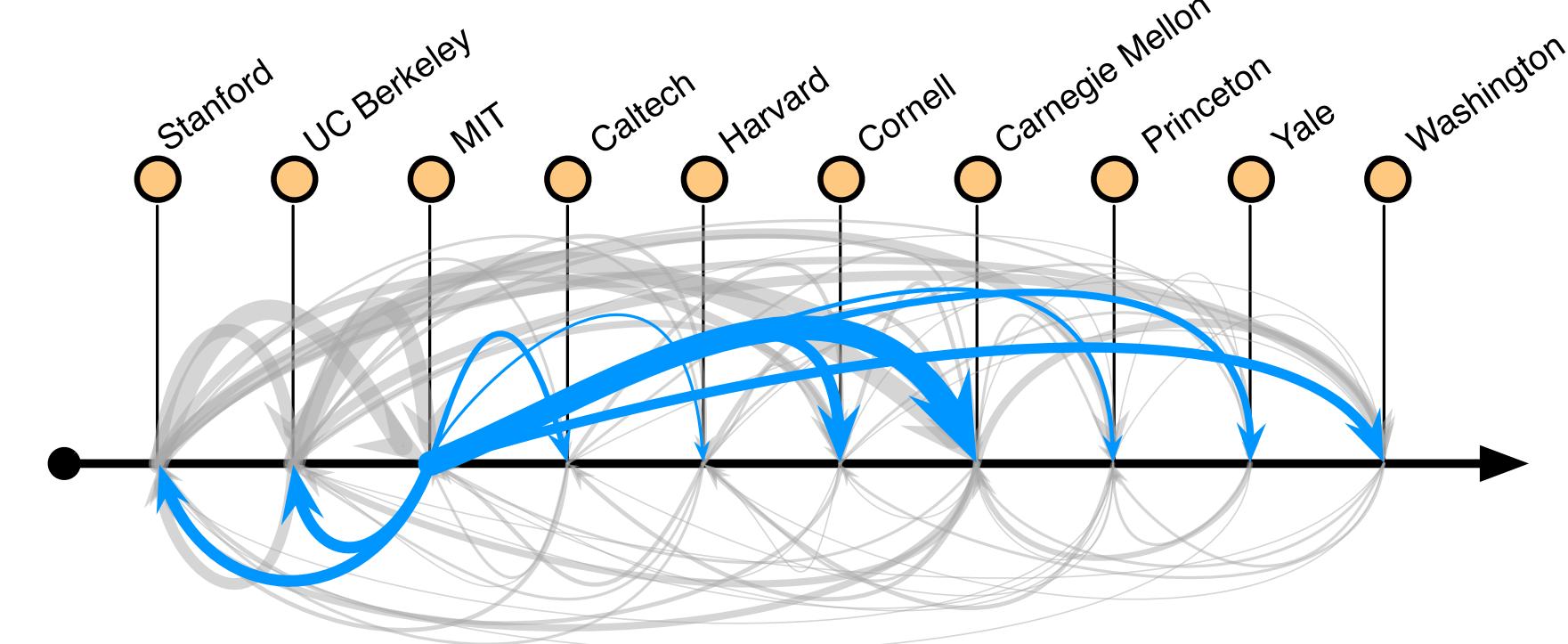
elite training does not drive scholarly productivity or prominence

working environment—not training—appears to drive scholarship

■ why do elite institutions dominate science?

doctoral prestige → faculty location

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



what drives productivity and prominence?

- ▶ prestige is a *structural variable* in the science of science
 - elite training does not drive scholarly productivity or prominence*
 - working environment—not training—appears to drive scholarship

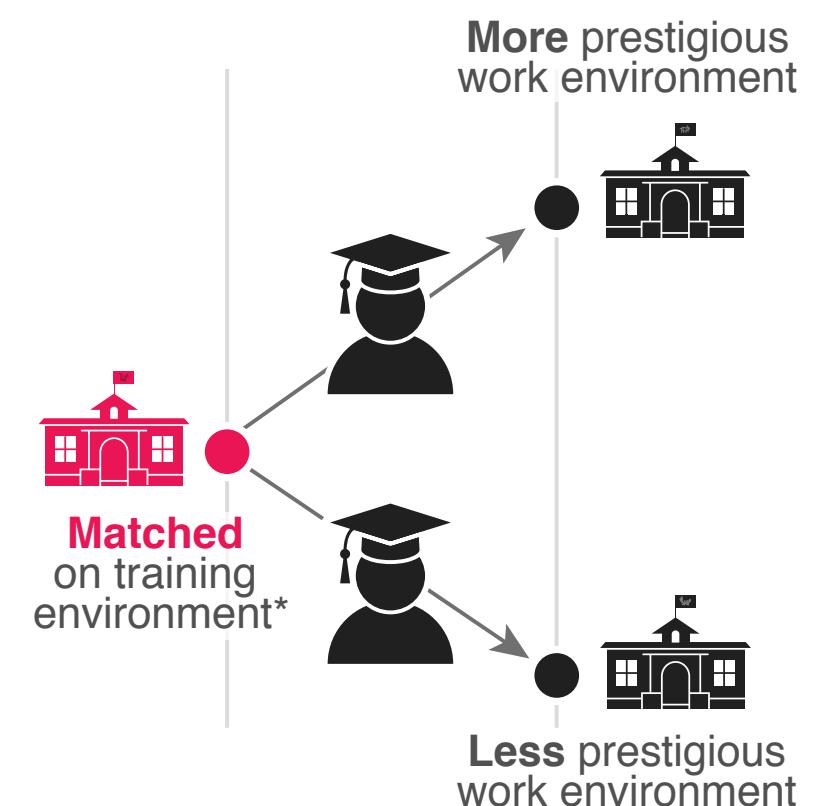
- why do elite institutions dominate science?

doctoral prestige → faculty location → scholarly impact

these results, Way et al. (2019)

- working environment causes higher impact (via 'unknown' mechanism)
- an *environmental* mechanism for the cumulative advantage of elite scientists

but how?



a labor advantage

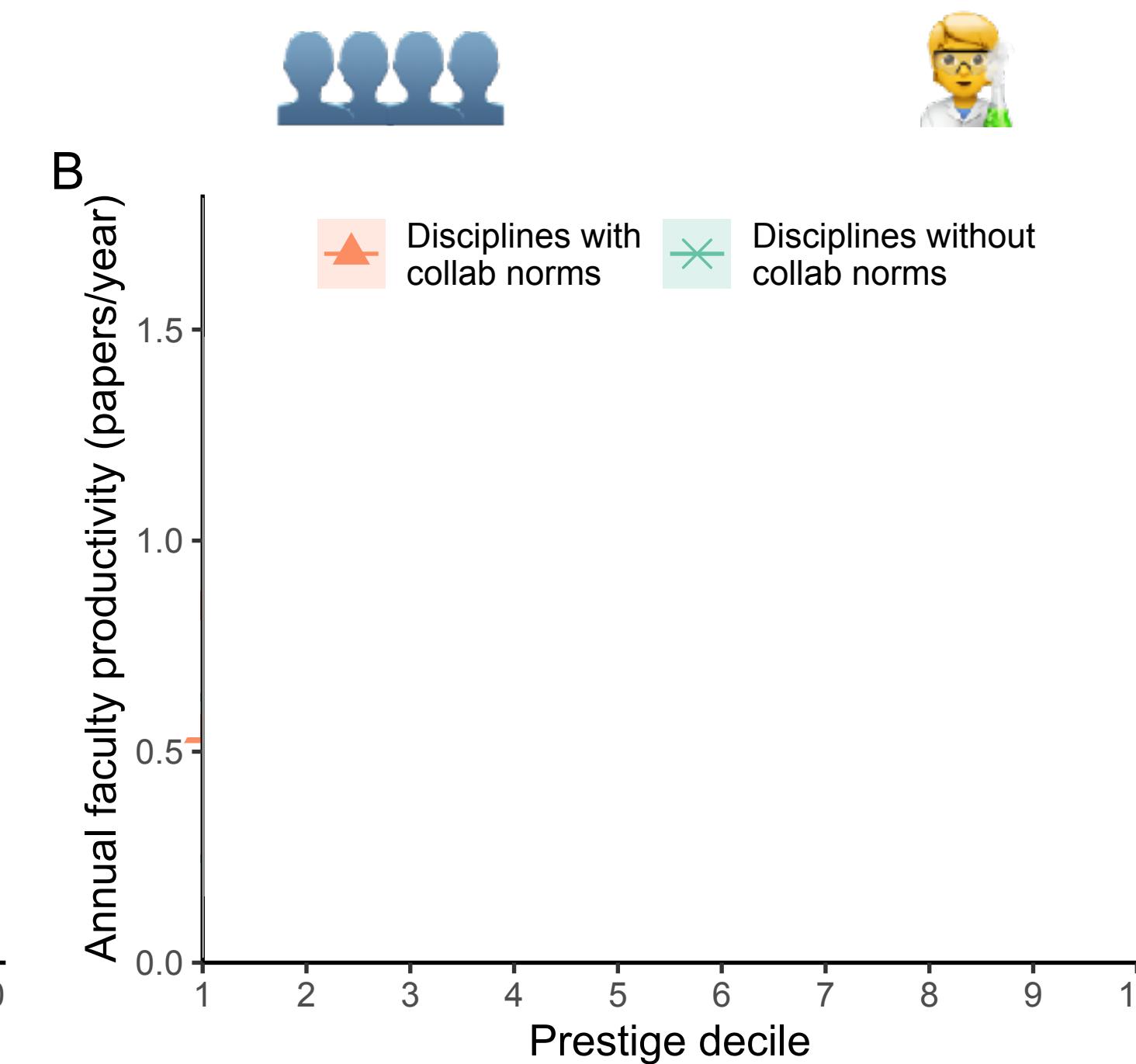
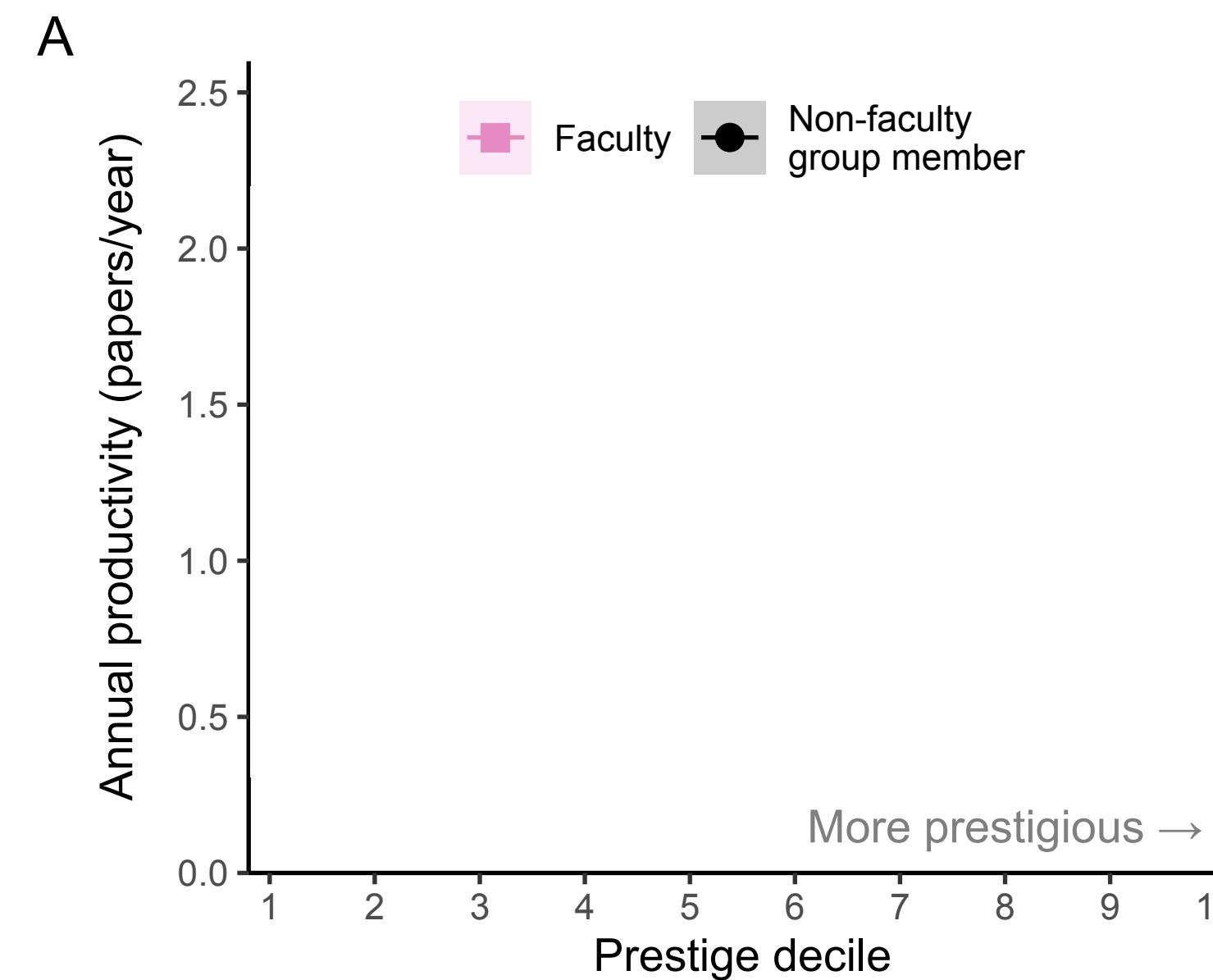
prestige → larger research groups → productivity

a labor advantage

prestige → larger research groups → productivity

linked data on 78,802 regular faculty at 262 PhD-granting institutions + complete publication data (Web of Science)

decompose total productivity = group prod. + individual prod.

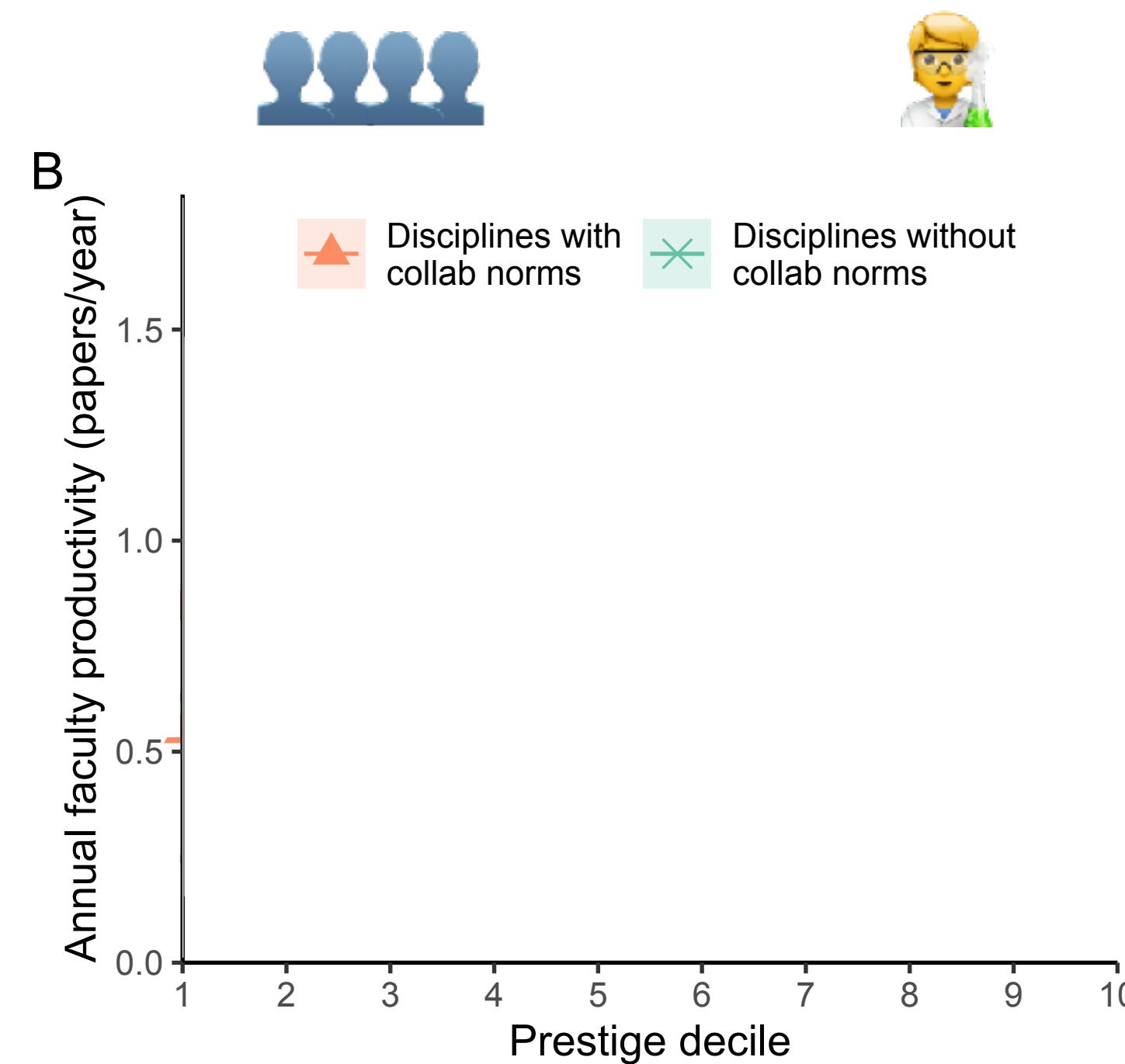
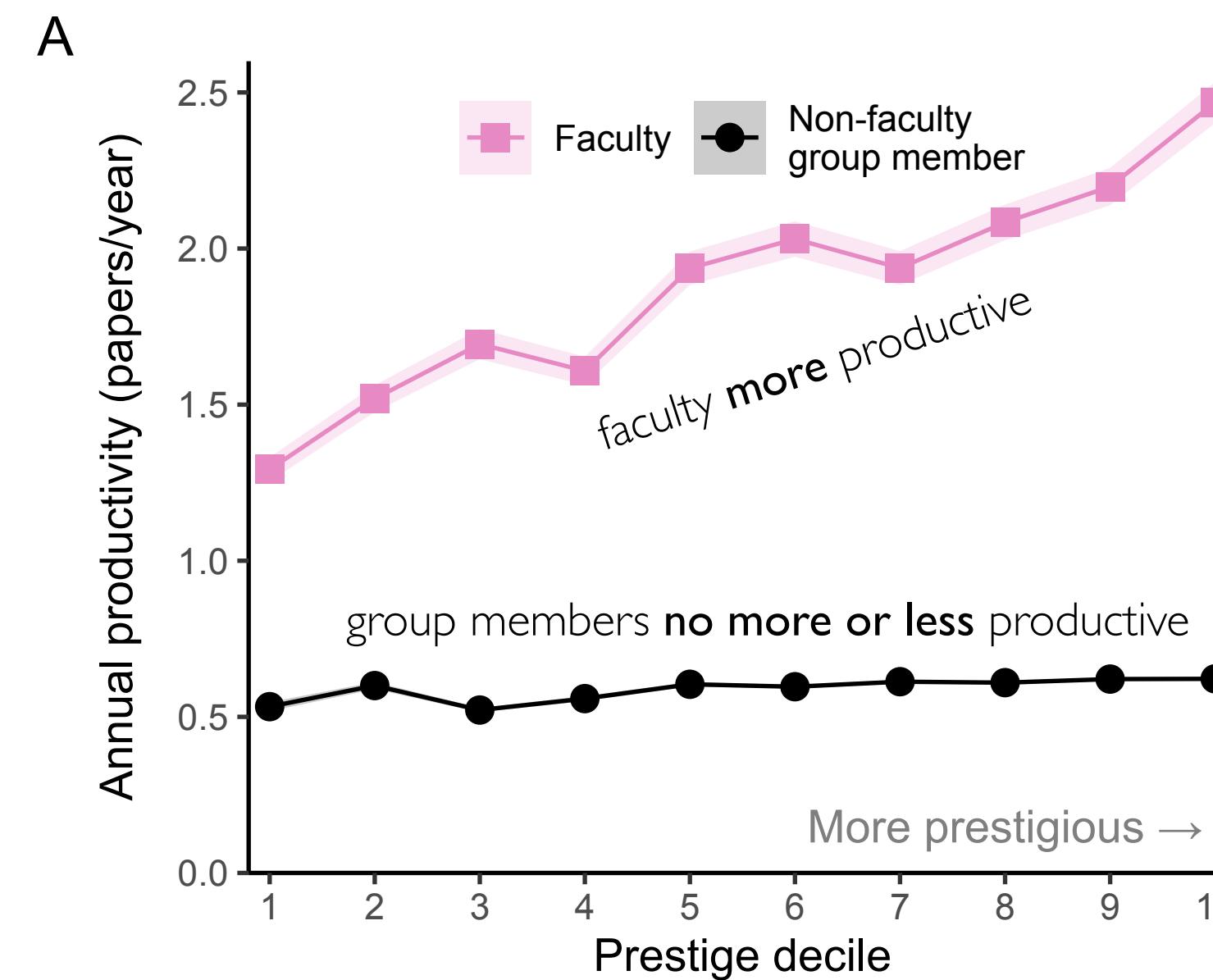


a labor advantage

prestige → larger research groups → productivity

linked data on 78,802 regular faculty at 262 PhD-granting institutions + complete publication data (Web of Science)

decompose total productivity = group prod. + individual prod.

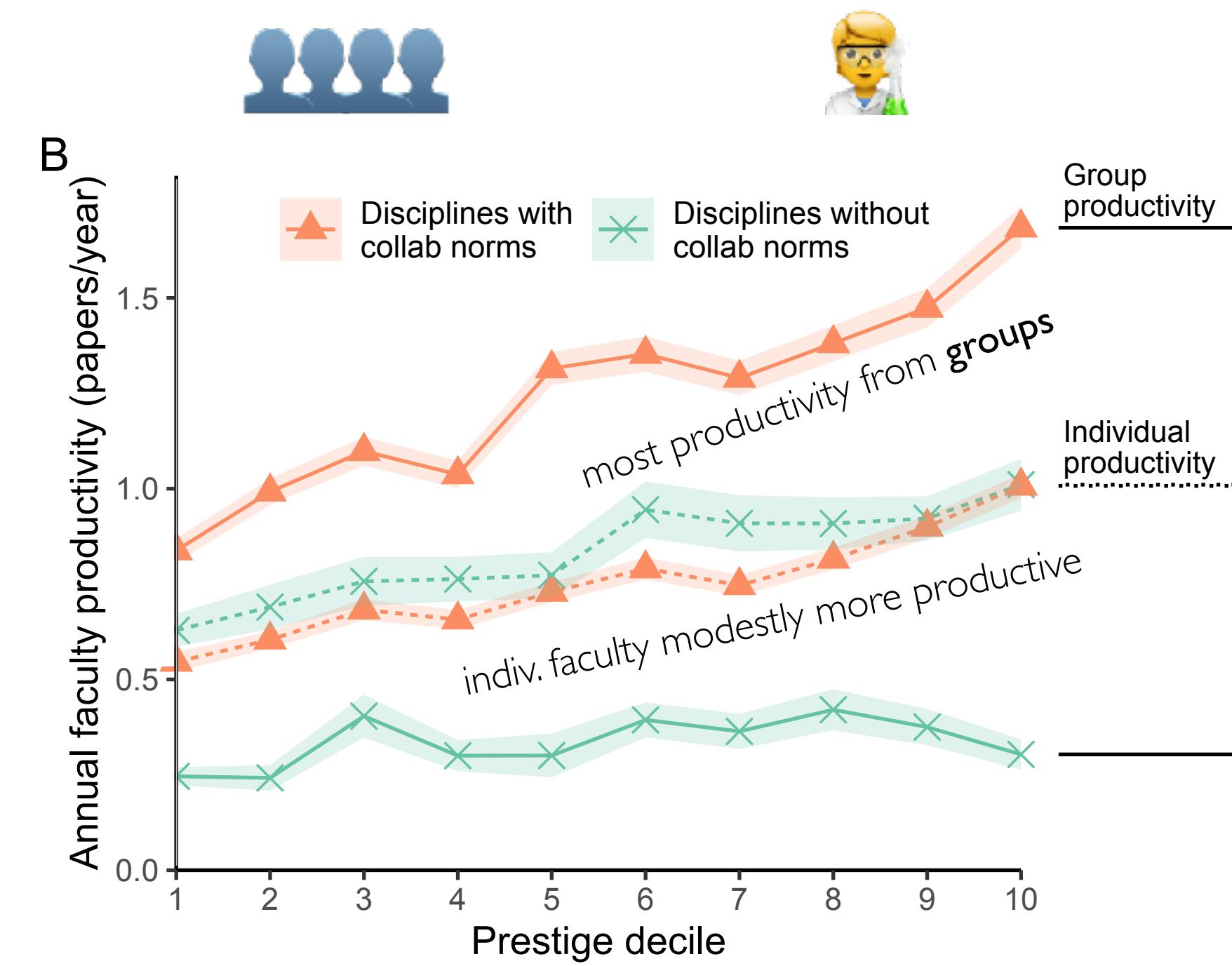
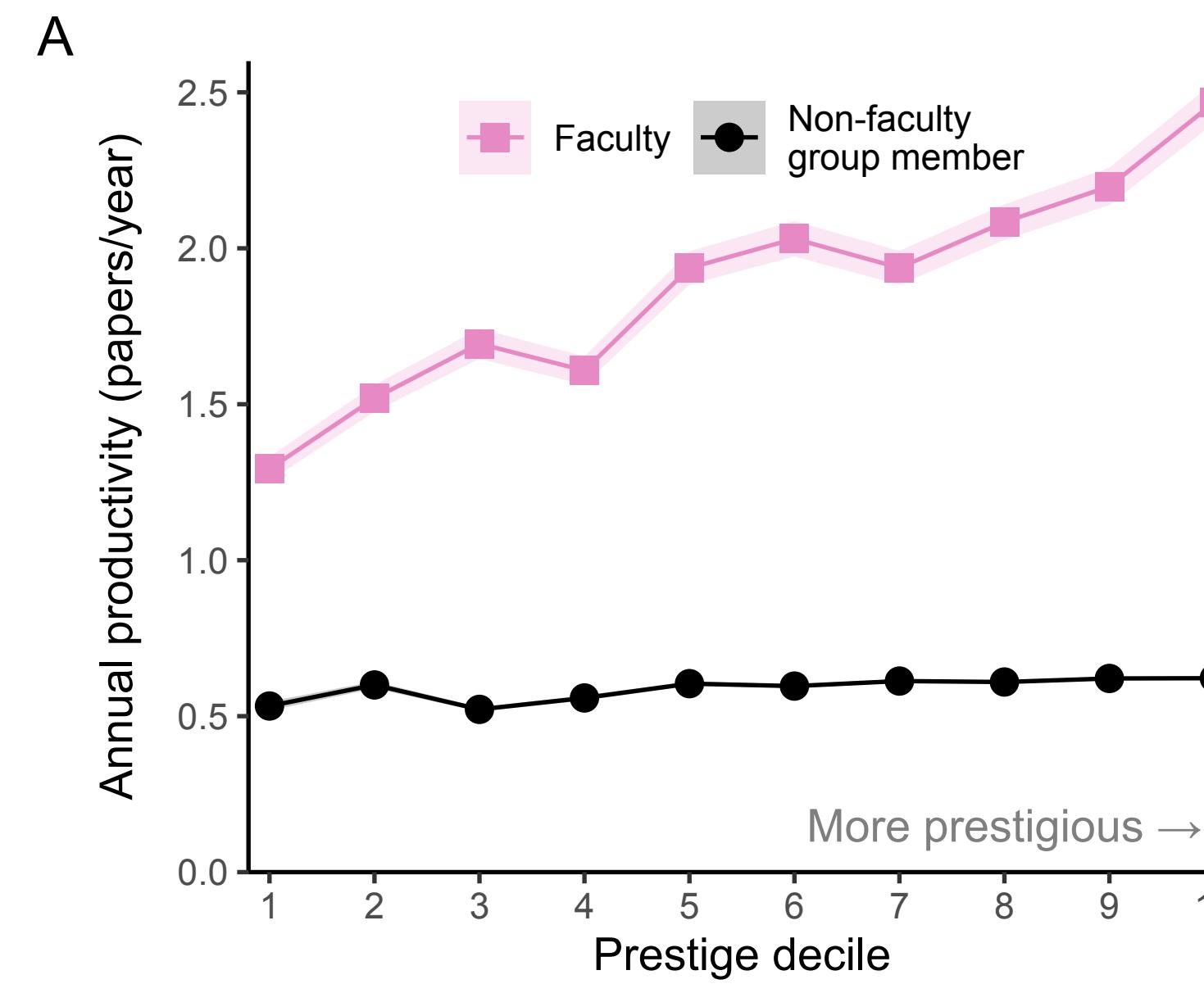


a labor advantage

prestige → larger research groups → productivity

linked data on 78,802 regular faculty at 262 PhD-granting institutions + complete publication data (Web of Science)

▶ group productivity growth explains most of prestige-productivity relationship

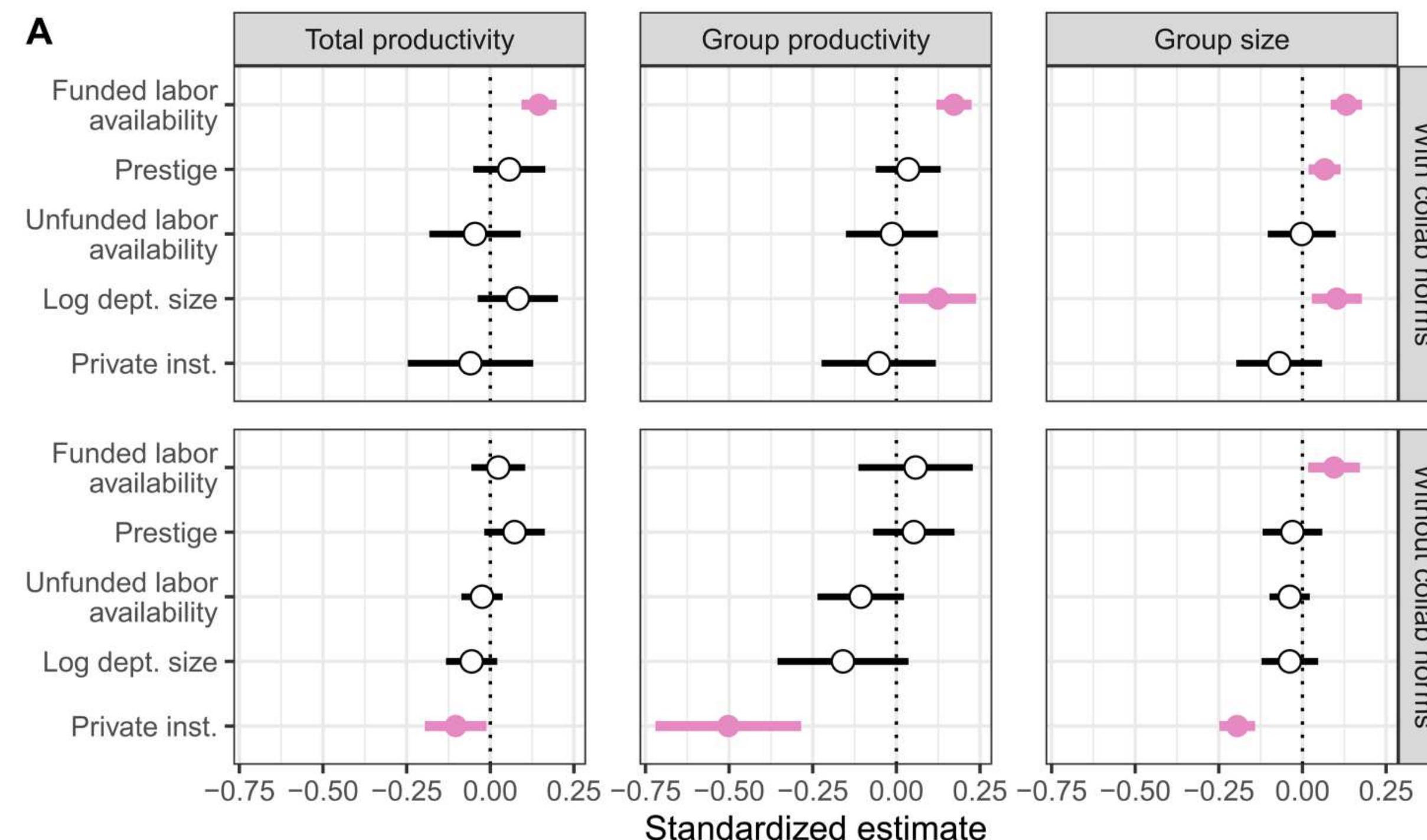


a labor advantage

prestige → larger research groups → productivity

linked data on 78,802 regular faculty at 262 PhD-granting institutions + complete publication data (Web of Science)

▶ and *labor* predicts total productivity & group productivity & group size



what drives productivity and prominence?

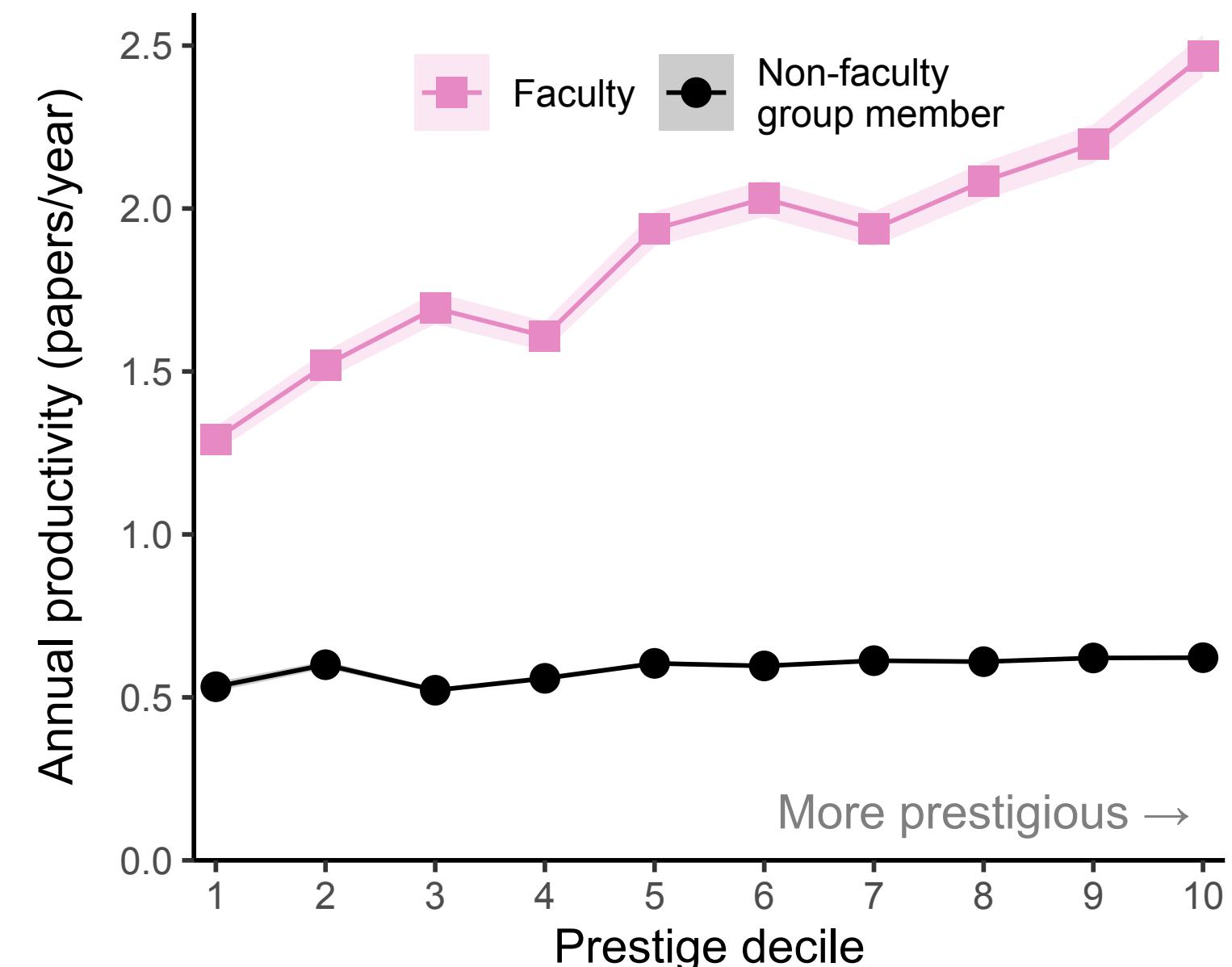
- ▶ prestige is a *structural variable* in the science of science
working environment drives scholarship, primarily via a *labor advantage*

- why do elite institutions dominate science?



- what else drives productivity and prominence?

networks...

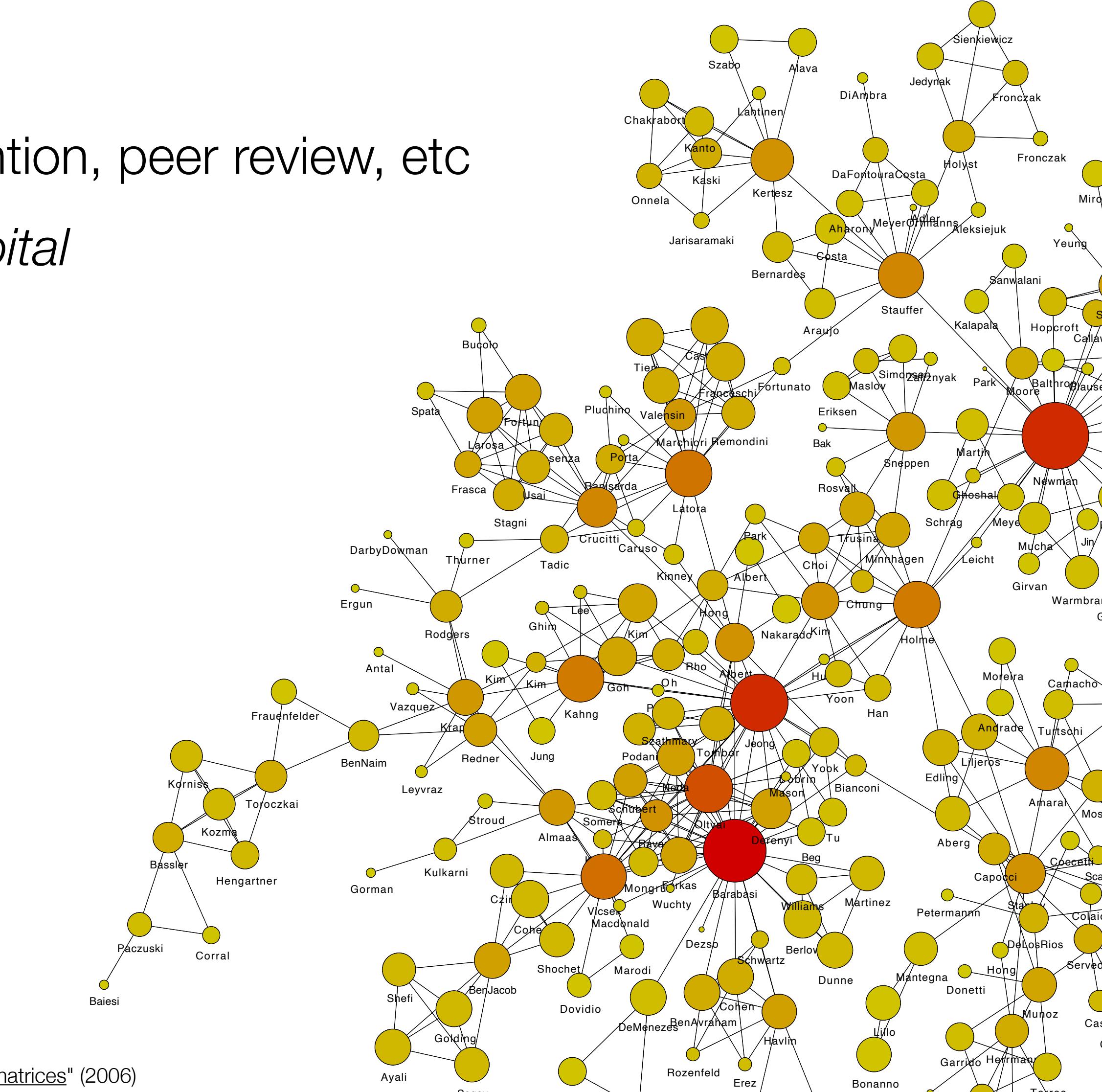


no scientist is an island unto themselves

▶ networks mediate most scientific activities:

scientific training, hiring, collaboration, teaching, attention, peer review, etc

they act like a form of *unequally distributed social capital*



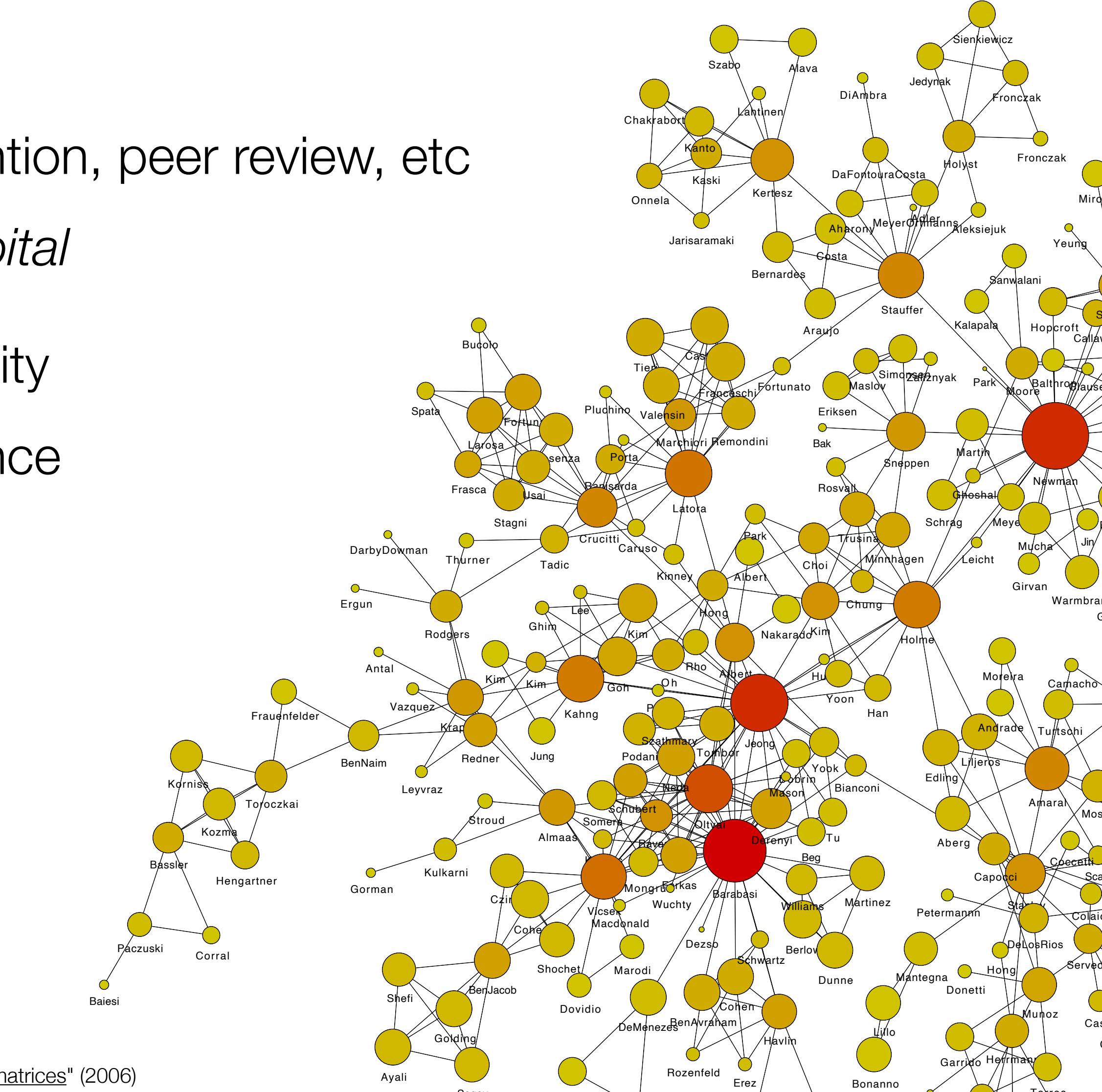
no scientist is an island unto themselves

▶ networks mediate most scientific activities:

scientific training, hiring, collaboration, teaching, attention, peer review, etc

they act like a form of *unequally distributed social capital*

- a productive collaborator → increases your productivity
- a prominent collaborator → increases your prominence



no scientist is an island unto themselves

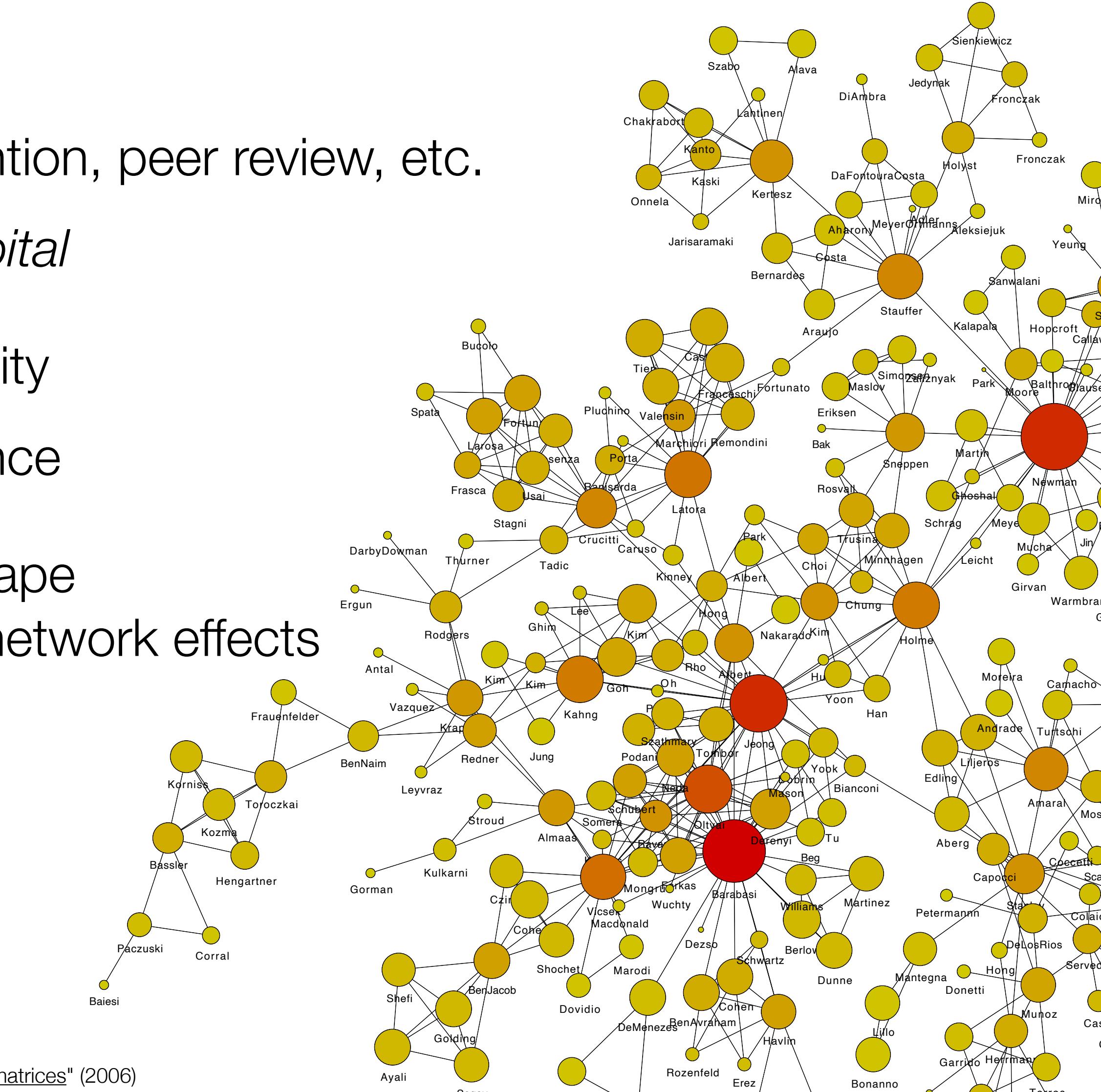
▶ networks mediate most scientific activities:

scientific training, hiring, collaboration, teaching, attention, peer review, etc.

they act like a form of *unequally distributed social capital*

- a productive collaborator → increases your productivity
- a prominent collaborator → increases your prominence

▶ understanding how gender, career age, and prestige shape productivity and prominence requires untangling these network effects

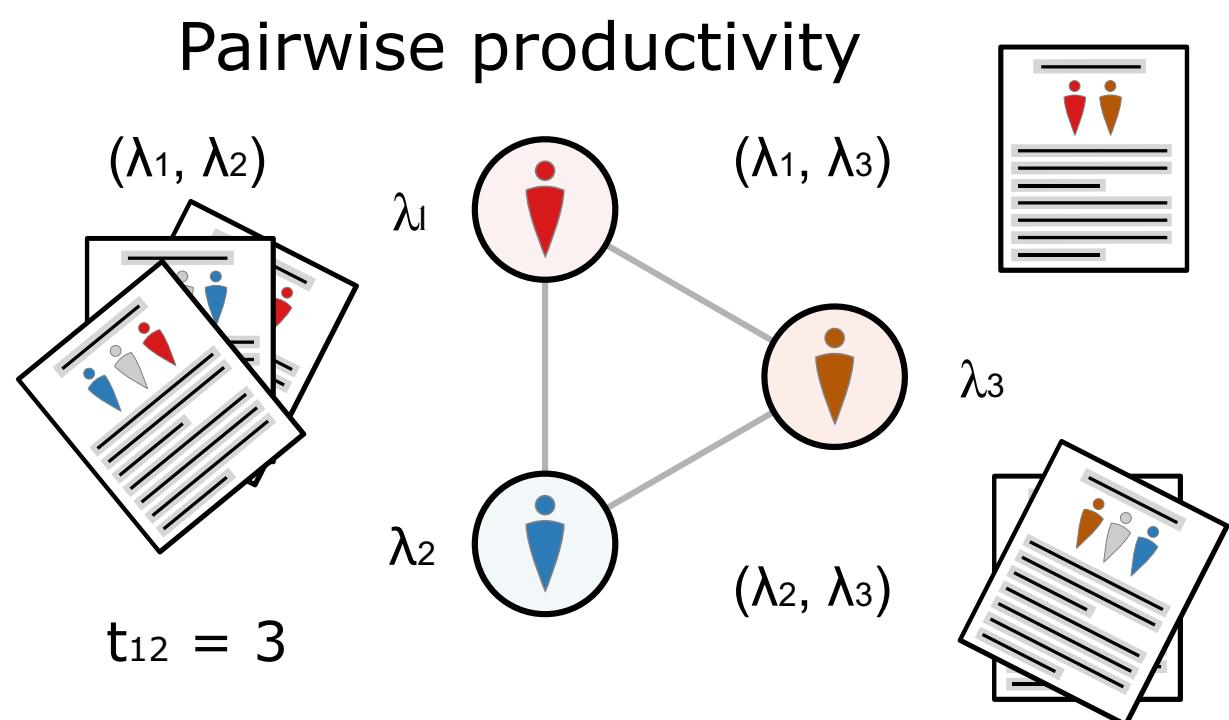


untangling the network effects

- ▶ generative model of pairwise *productivity* and of pairwise *prominence* → infer *individual* parameters
- applied to a large collaboration network, we can untangle collaborative effects
- estimate individual propensities for productivity and prominence
- ask how other career aspects covary with these propensities

untangling the network effects

- ▶ generative model of pairwise *productivity* and of pairwise *prominence* → infer *individual* parameters
- assume: joint *productivity* of i and j is linear function $\lambda_i + \lambda_j$ of their latent individual productivities



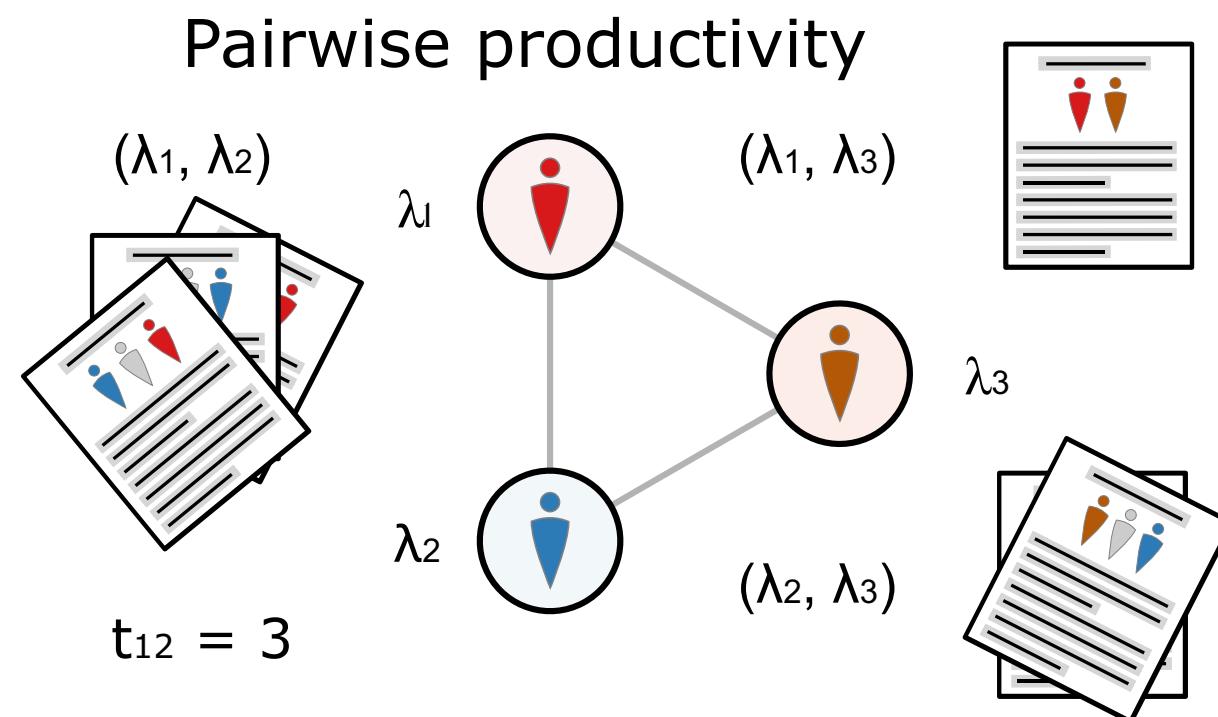
number (i, j) -coauthored papers

$$\Pr(N_{ij}, t_{ij} | \lambda_i, \lambda_j) = \text{Poisson}([\lambda_i + \lambda_j]t_{ij})$$

given time period

untangling the network effects

- ▶ generative model of pairwise *productivity* and of pairwise *prominence* → infer *individual* parameters
 - assume: joint *productivity* of i and j is linear function $\lambda_i + \lambda_j$ of their latent individual productivities

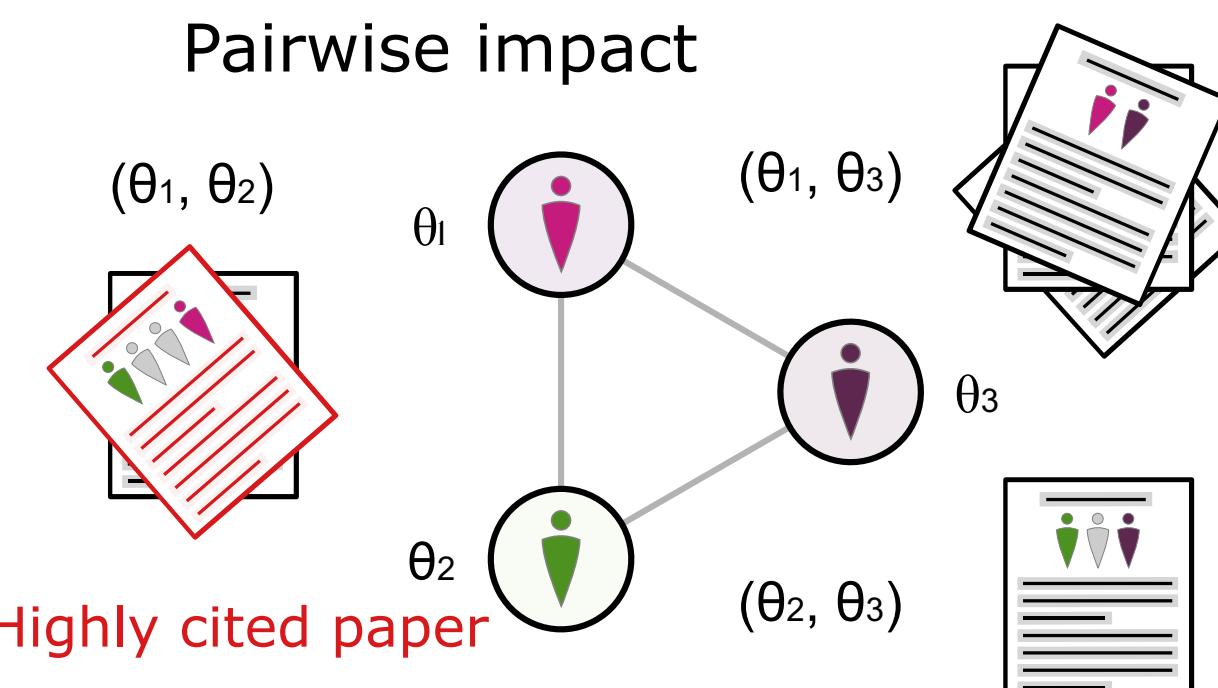


number (i, j)-coauthored papers

$\Pr(N_{ij}, t_{ij} | \lambda_i, \lambda_j) = \text{Poisson}([\lambda_i + \lambda_j]t_{ij})$

given time period

- assume: joint *prominence* is linear function $\theta_i + \theta_j$ of their latent individual impacts



number (i, j)-coauthored papers

$\Pr(N_{ij}, m_{ij} | \theta_i, \theta_j) = \text{Binomial}(N_{ij}, [\theta_i + \theta_j])$

number of "high impact" papers

estimating from data

Microsoft Academic Graph (MAG) 1950–2019



- select 6 STEM fields (biology, chemistry, CS, math, medicine, and physics)
- focus on first-last author pairs — mitigates middle-author effects
- focus on mid-career researchers — 10+ papers by 15th year of publishing history
- define 'highly cited' — in upper 8% for given year-field

Institutional prestige

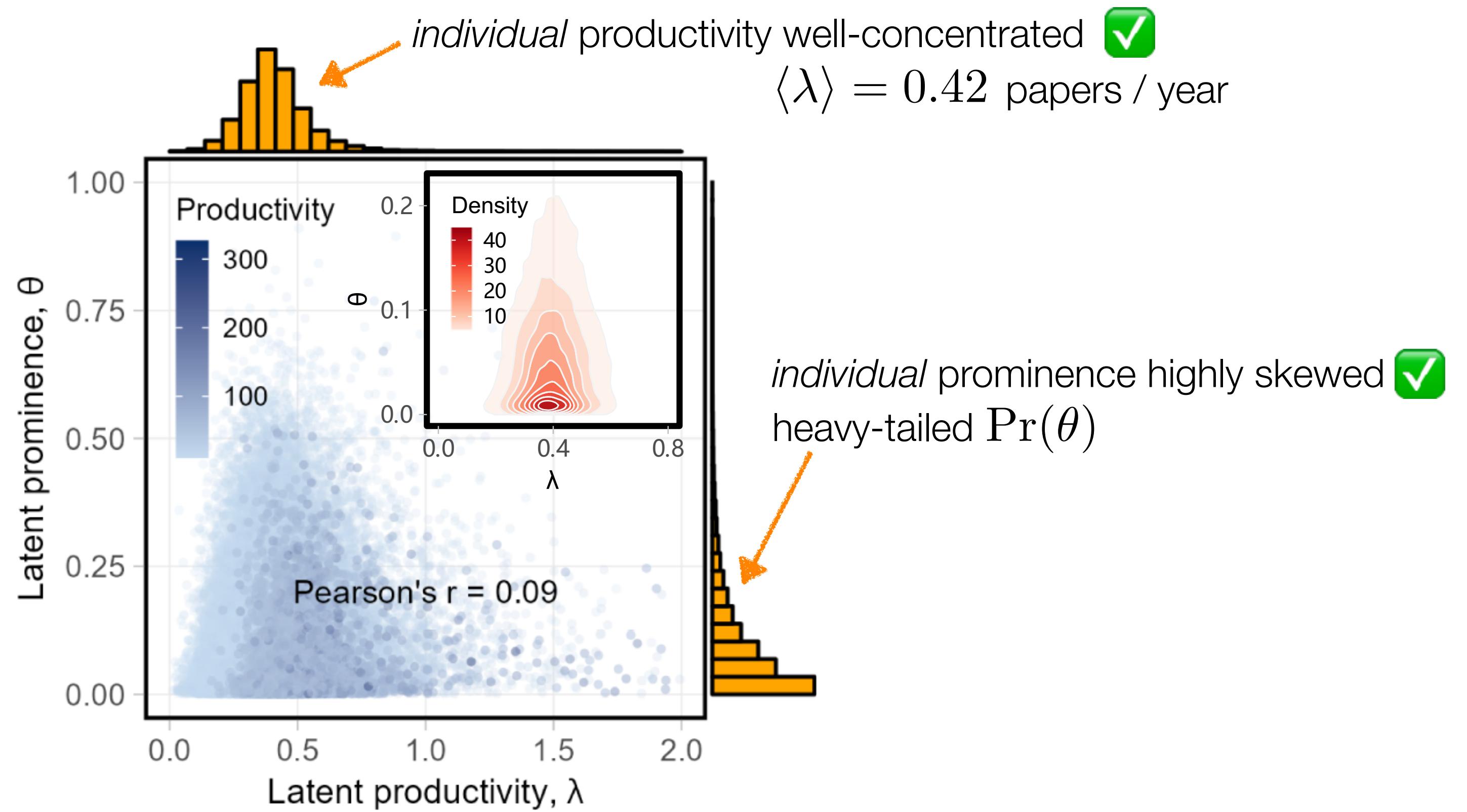
- "elite" — top 10 institutions under z -score of total highly cited papers over time period

Estimating parameters

- estimate using network 1950– T , for variable T in [1975,2017]
- bootstrapped convex optimization, then record mean λ_i and θ_i

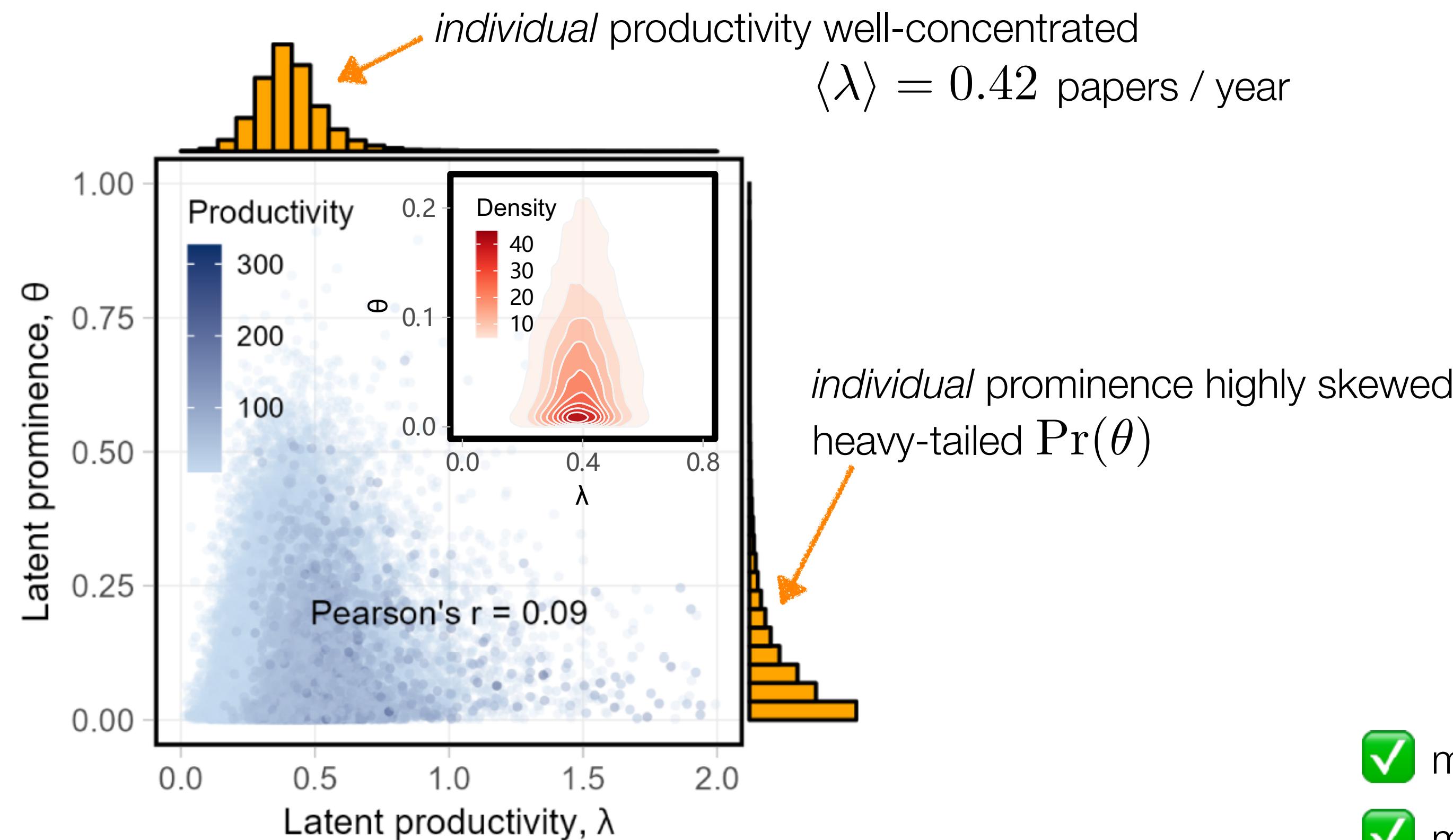
model checking

▶ applied to 198,202 mid-career STEM researchers 1975-2017



model checking

▶ applied to 198,202 mid-career STEM researchers 1975-2017



individual prominence highly skewed
heavy-tailed $\text{Pr}(\theta)$

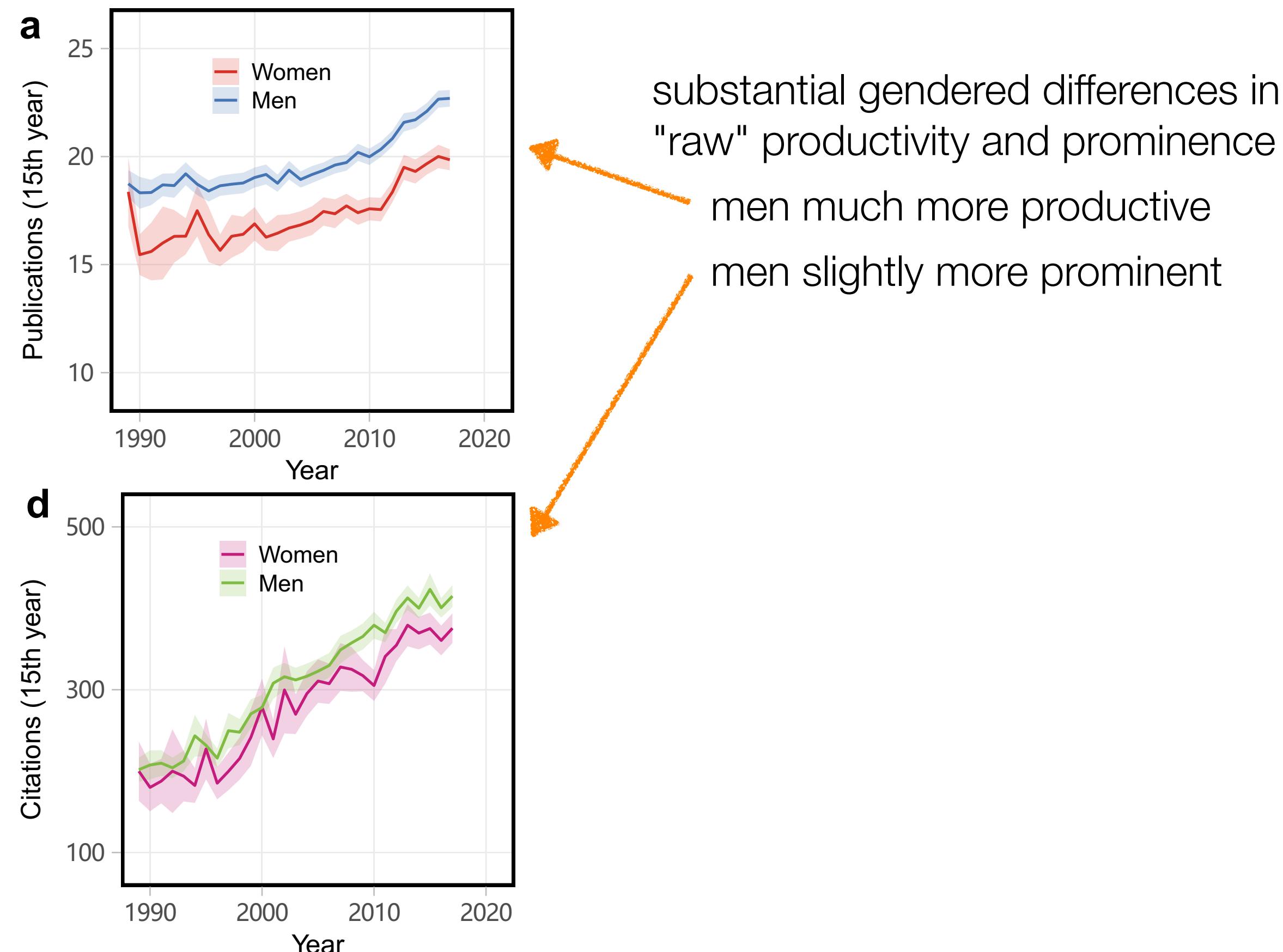
	Prestige	Papers	Citations	λ	θ	High λ coauthors	High θ coauthors
Prestige		0.06	0.15	0.02	0.15	0.04	0.13
Papers	0.06		0.4	0.21	-0.02	0.7	0.44
Citations	0.15	0.4		0.12	0.38	0.27	0.49
λ	0.02	0.21	0.12		0.15	0.31	0.14
θ	0.15	-0.02	0.38	0.15		0.06	0.25
High λ coauthors	0.04	0.7	0.27	0.31	0.06		0.43
High θ coauthors	0.13	0.44	0.49	0.14	0.25	0.43	

- ✓ my paper count = highly correlated with high λ coauthors
- ✓ my citations = well correlated with high λ & θ coauthors

gender vs. productivity & prominence

▶ applied to 198,202 mid-career STEM researchers 1975-2017

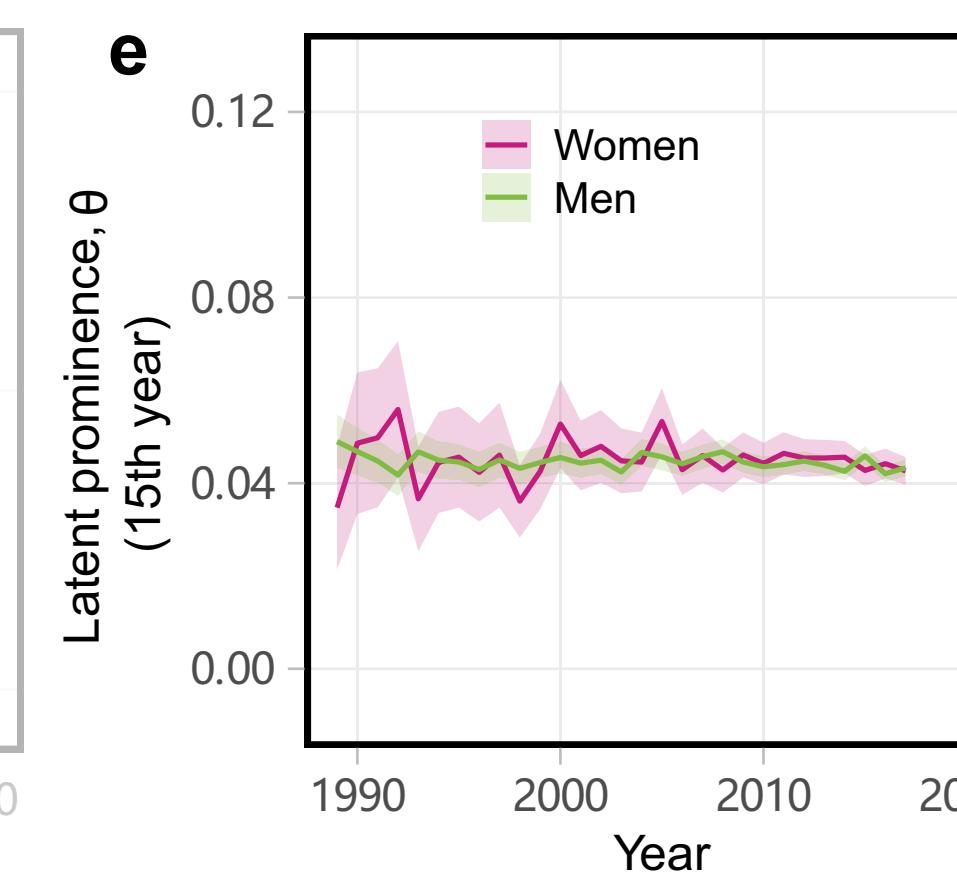
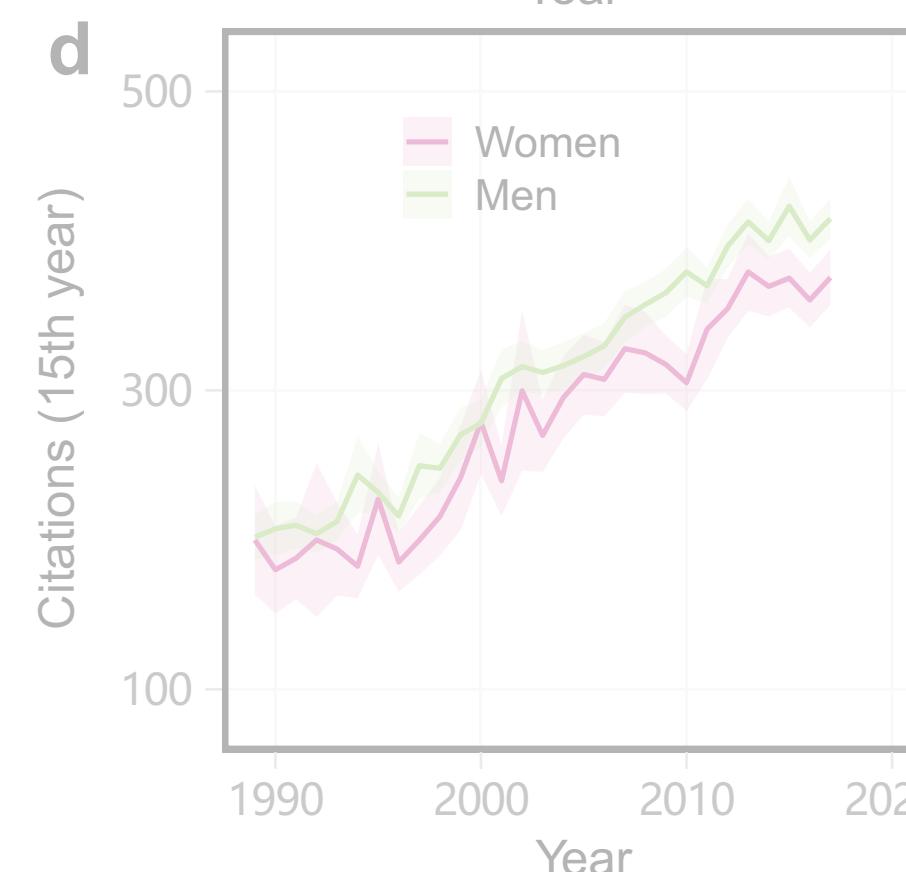
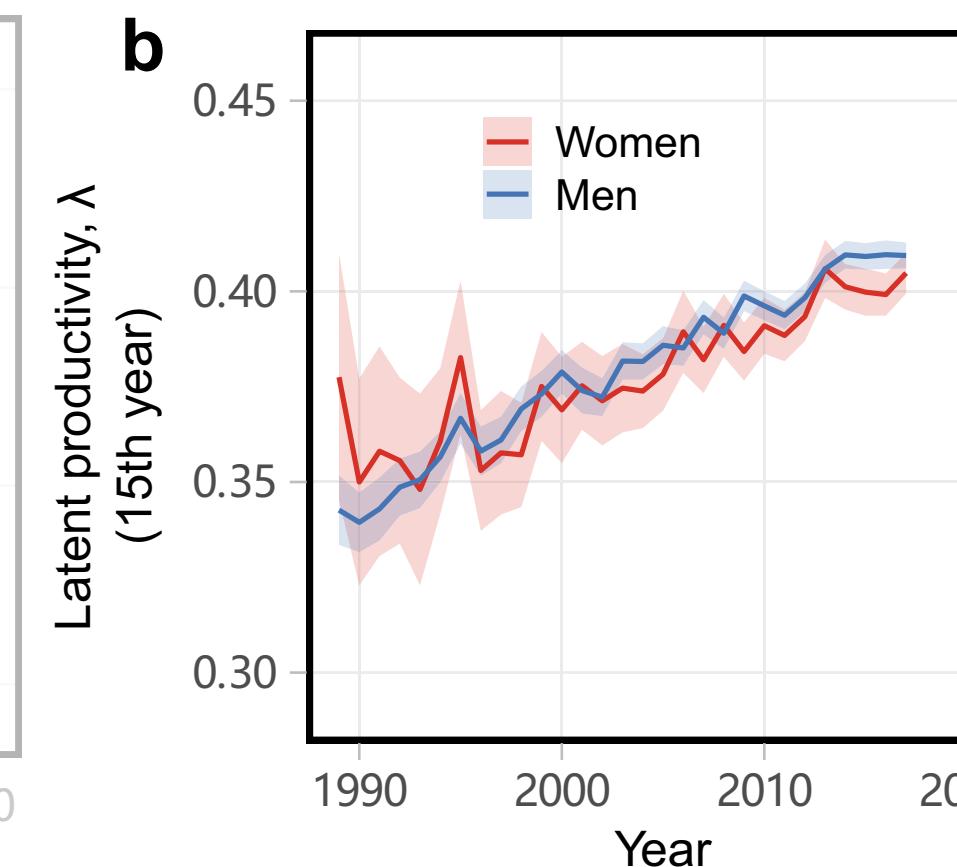
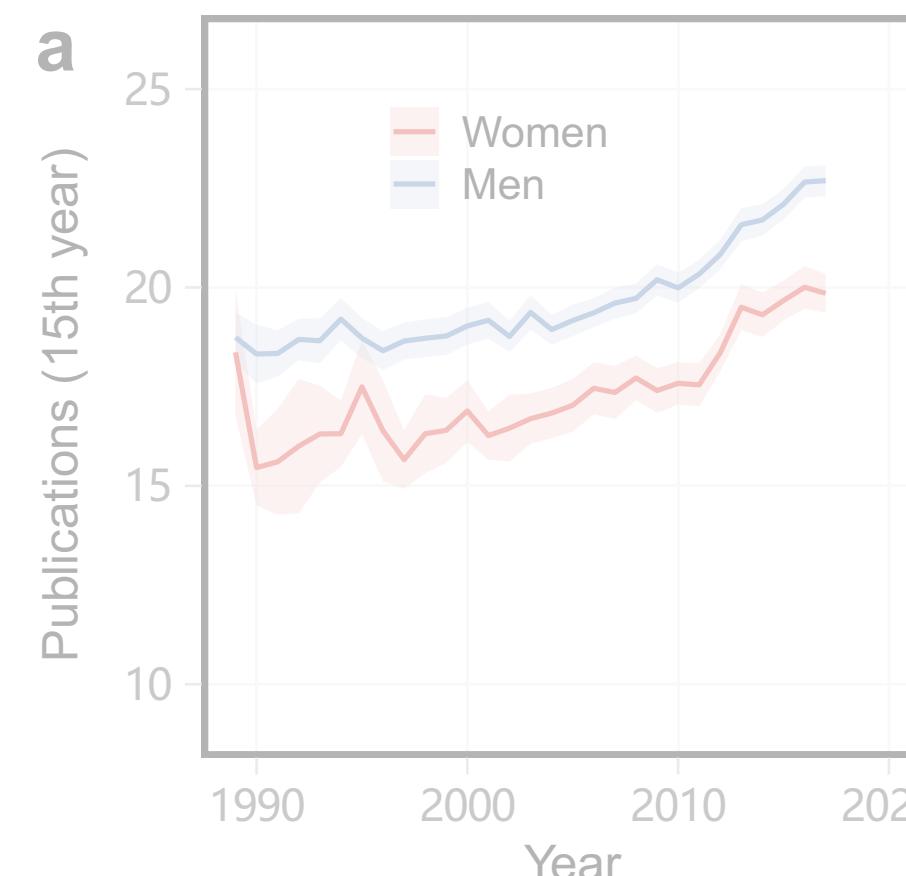
- compare men and women across time



gender vs. productivity & prominence

▶ applied to 198,202 mid-career STEM researchers 1975-2017

■ compare men and women across time → their networks are different



but: latent productivity and prominence
is *not* gendered

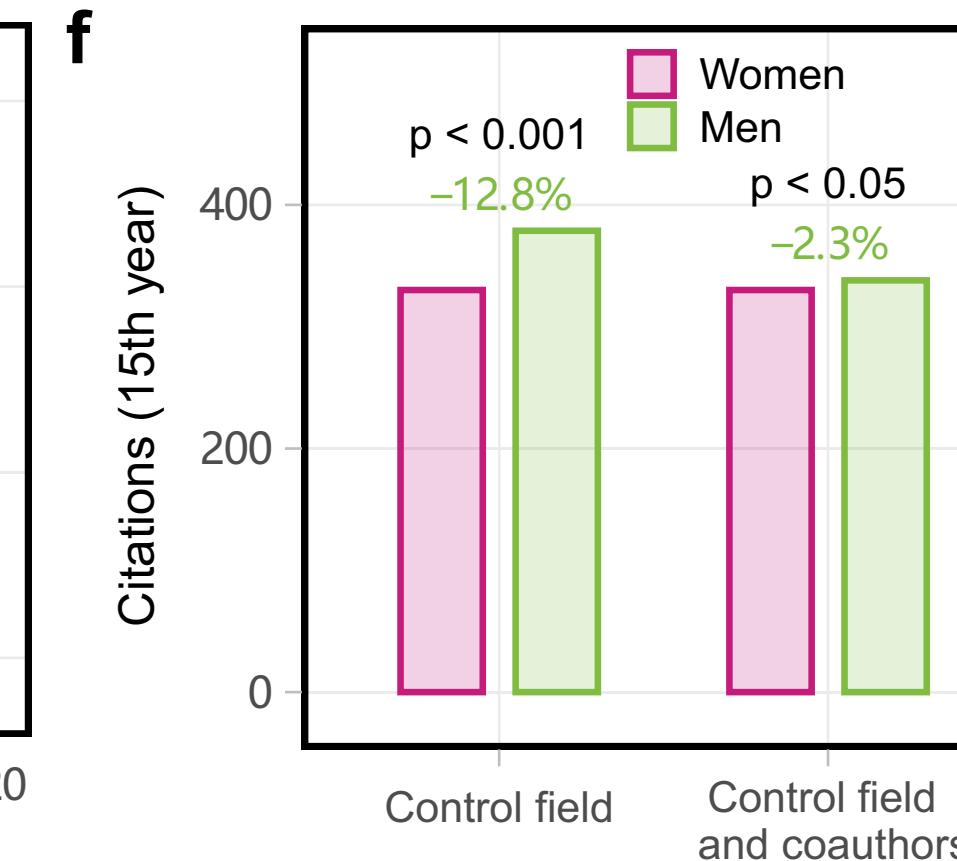
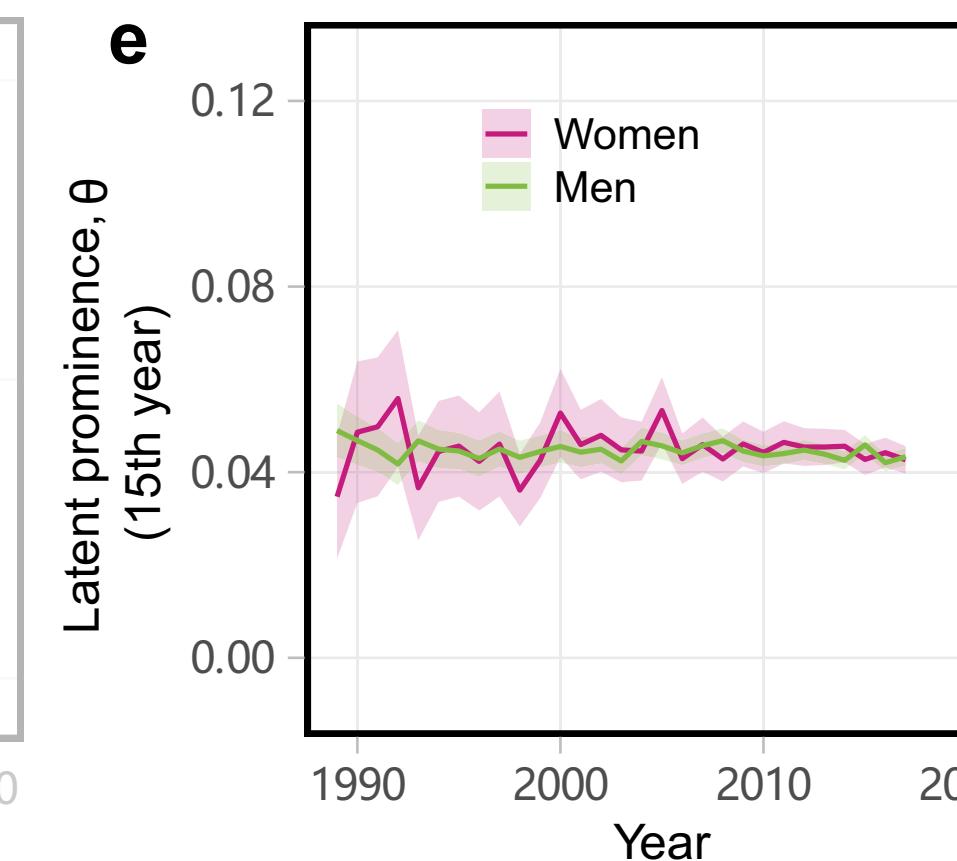
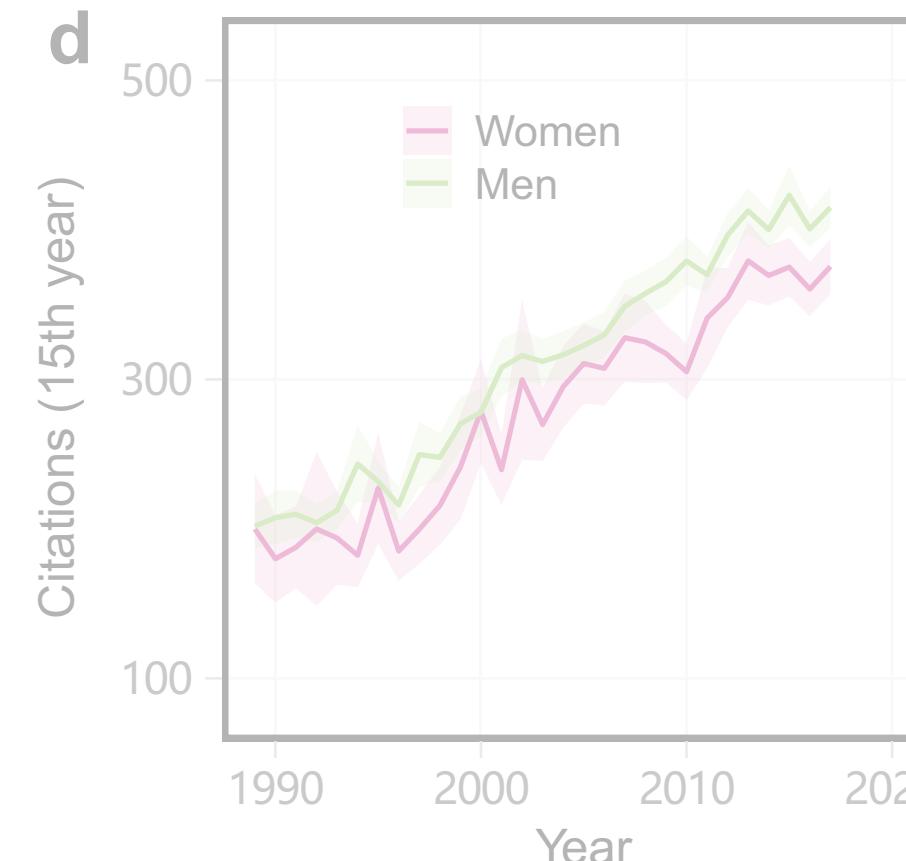
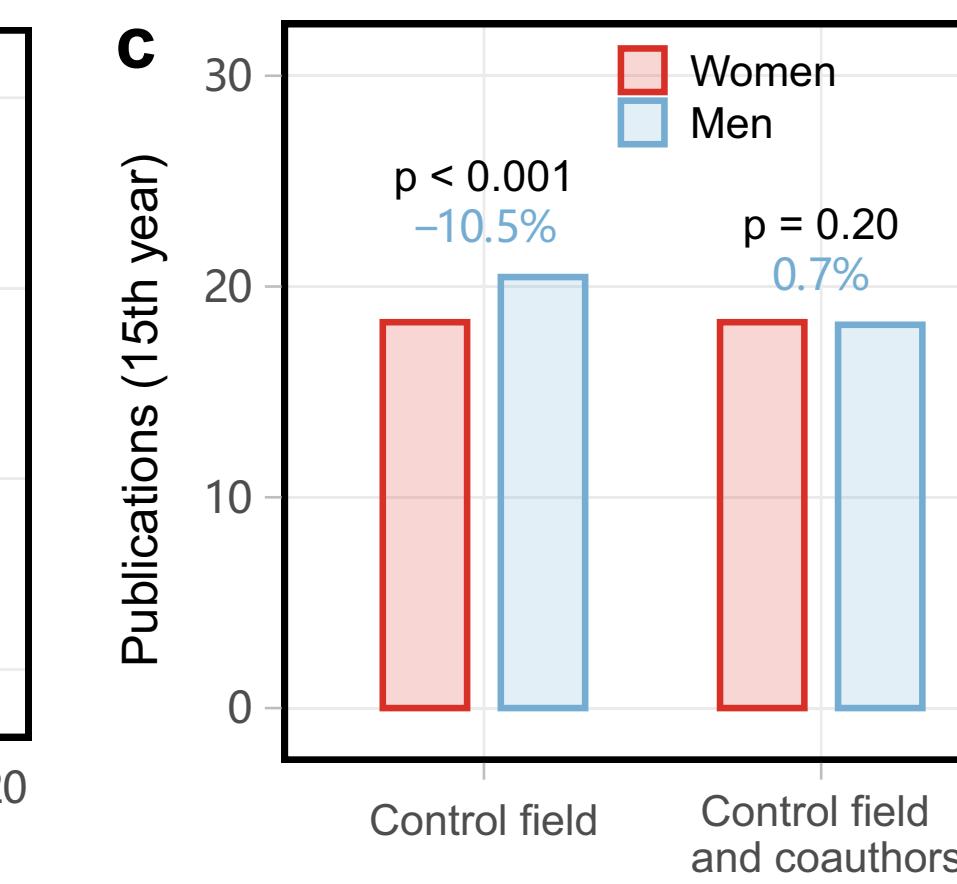
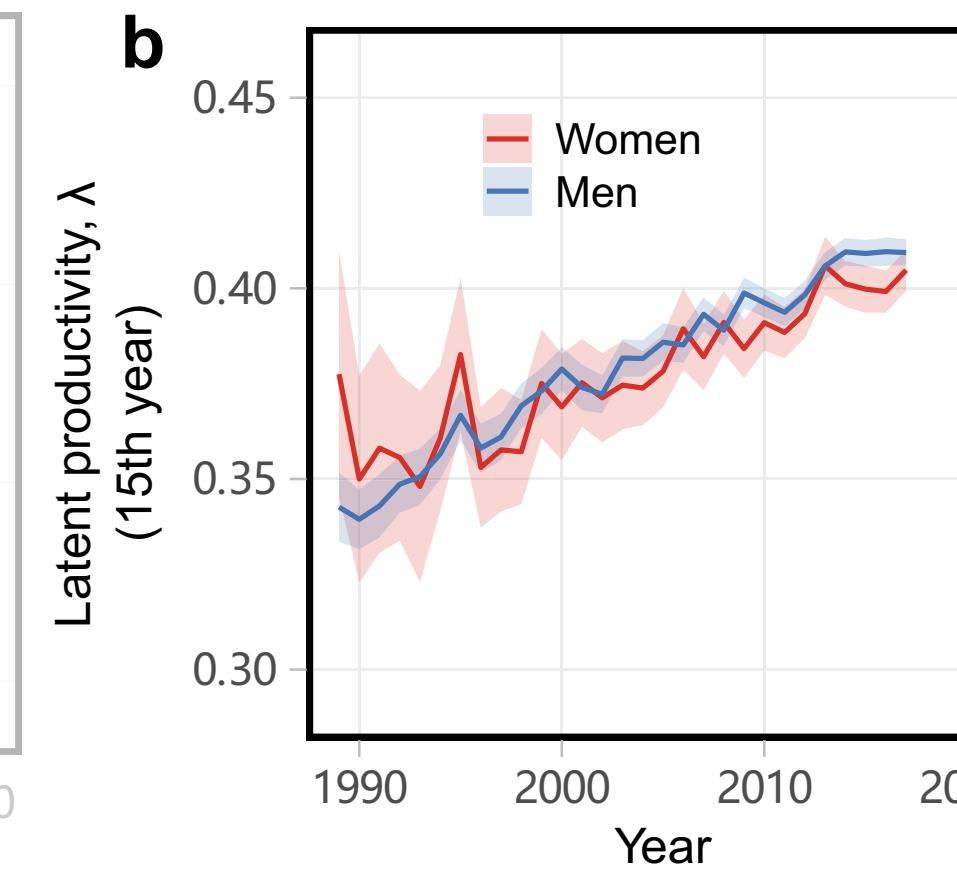
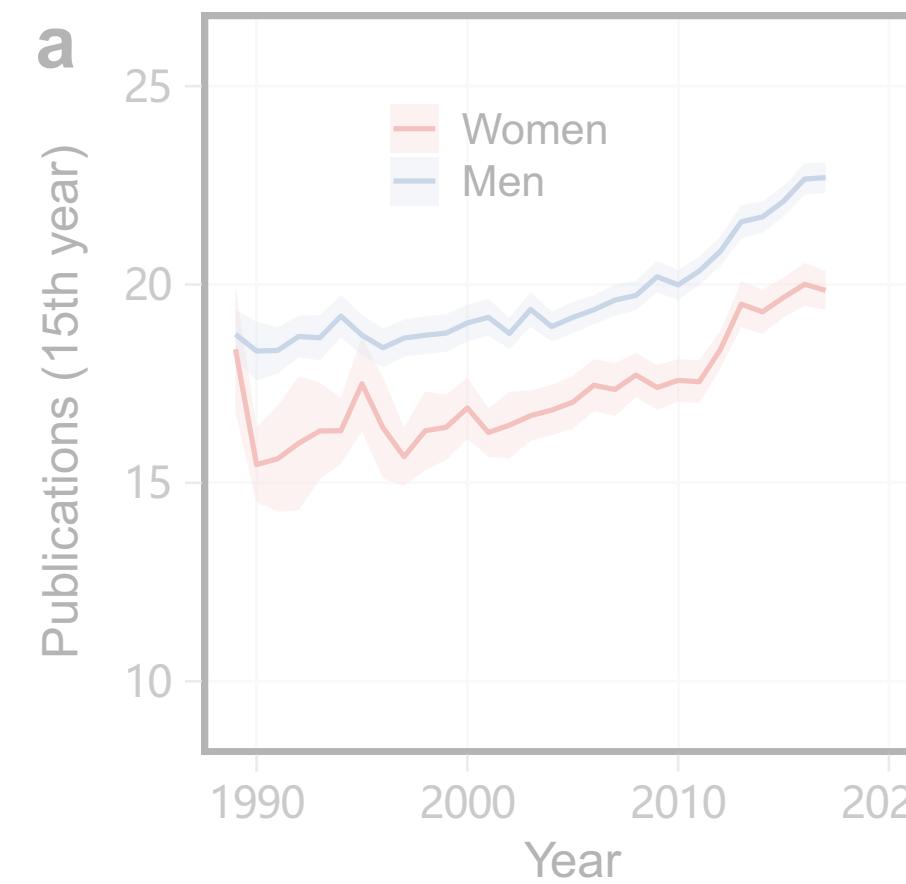
- ▶ size and composition of collaboration networks is gendered
- ▶ latent productivities increase steadily
- ▶ latent prominence stable over time

*not causal, but implies effects of known gendered causal factors on productivity (eg, parenthood) may operate by reshaping collaborating networks

gender vs. productivity & prominence

▶ applied to 198,202 mid-career STEM researchers 1975-2017

- compare men and women across time → their networks are different

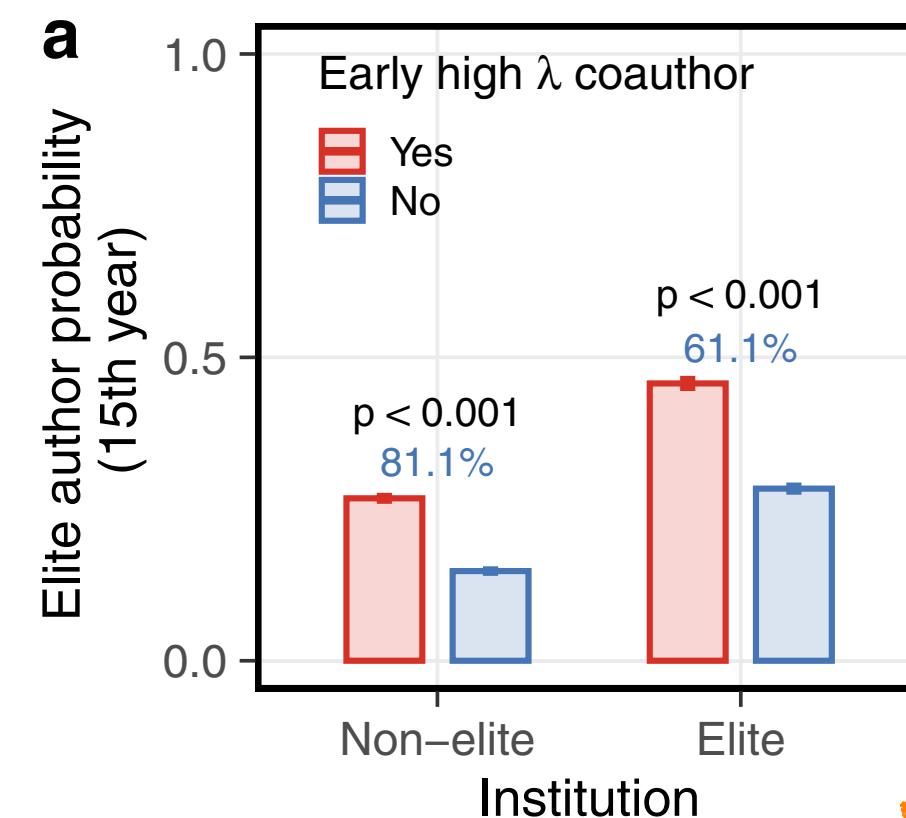


matching women/men on institutional prestige + age + field OR field+num. coauthors:
collaboration rates (num. coauthors) can explain gendered differences in "raw" productivity and prominence ✓

role of elite collaborators

▶ how does an early-career collaboration with an elite senior researcher influence prominence?

- elite senior researches with high λ or high θ



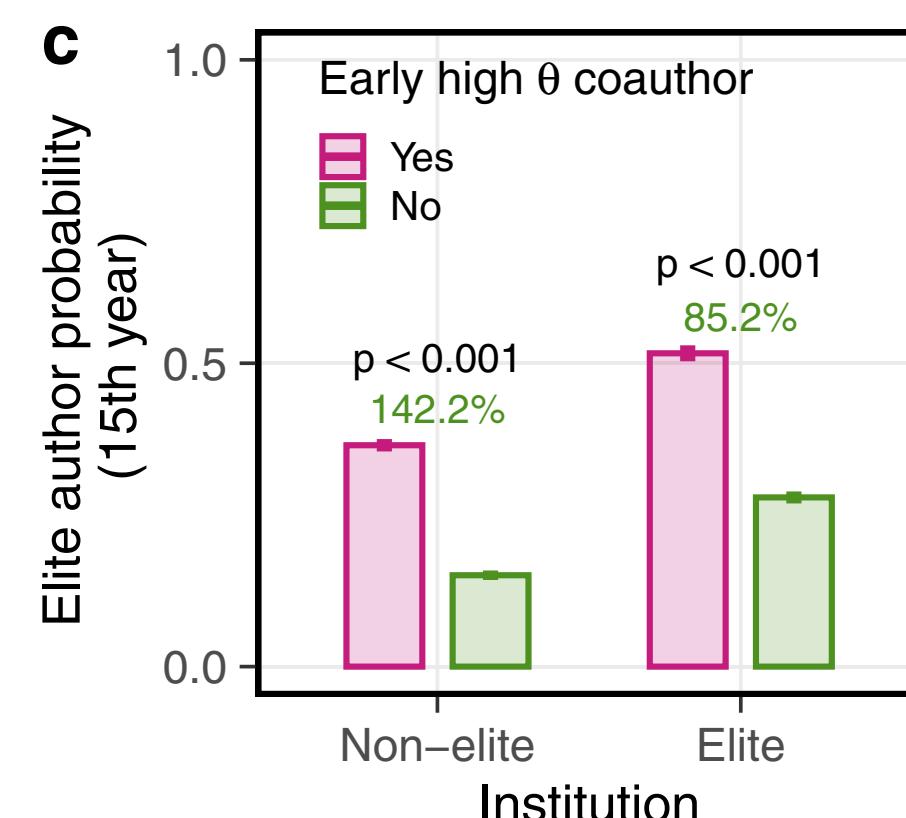
high- λ or early high- θ early collaborator substantially increases likelihood of high prominence in mid-career

▶ common at elite institutions

▶ effect appears at non-elite institutions too ✓

Pr. of high- λ early collab = 0.18 (elite) & 0.15 (non-elite)

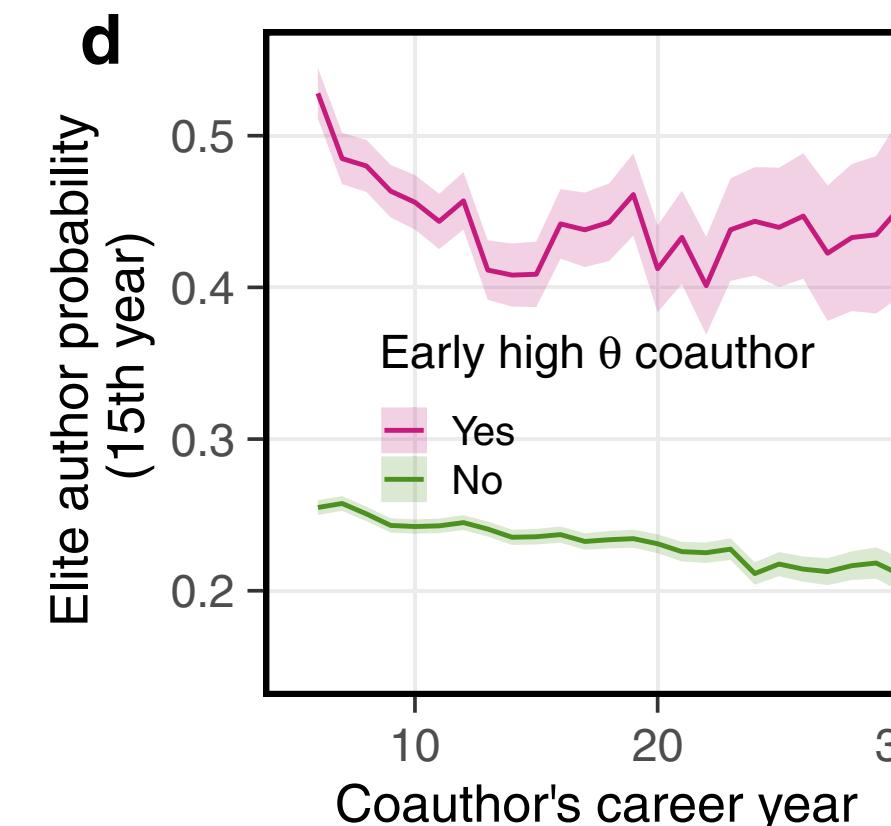
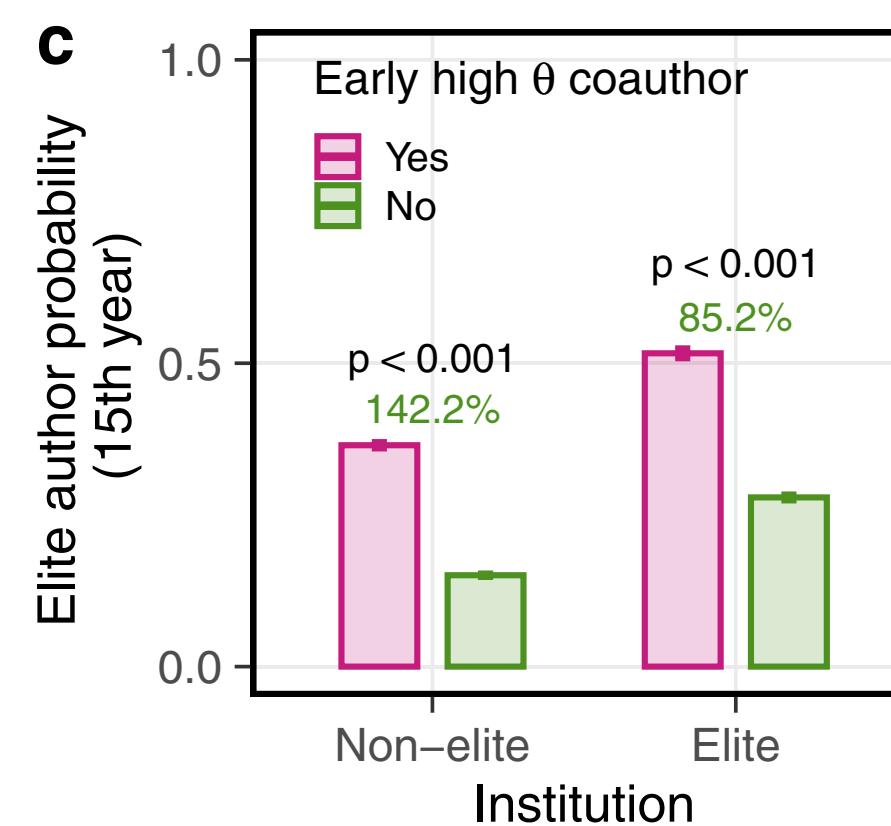
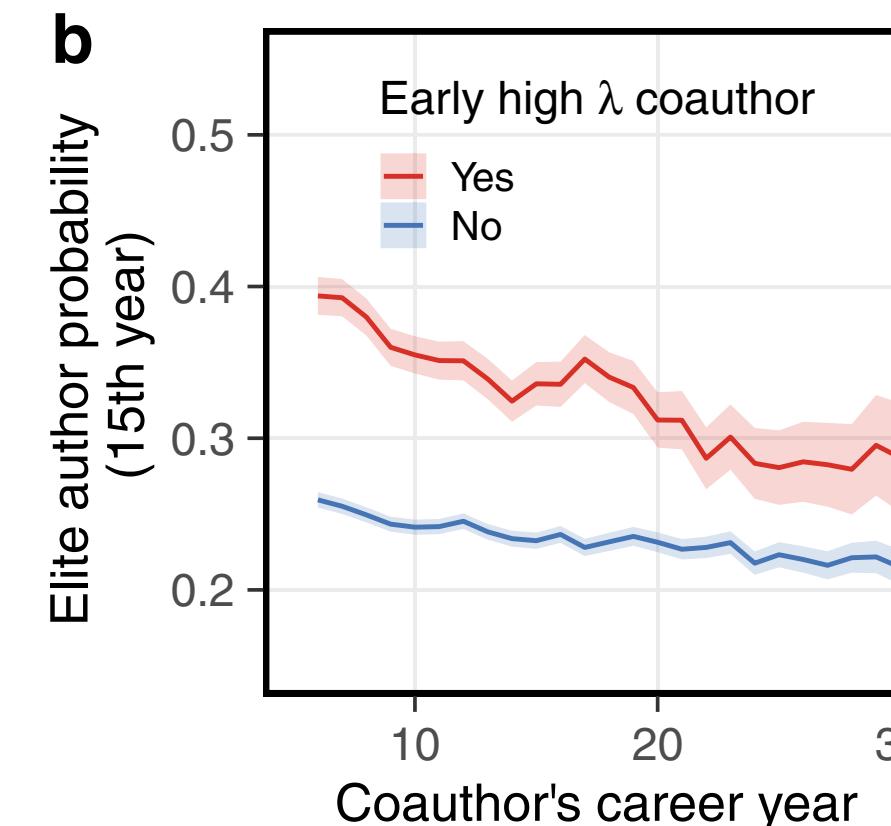
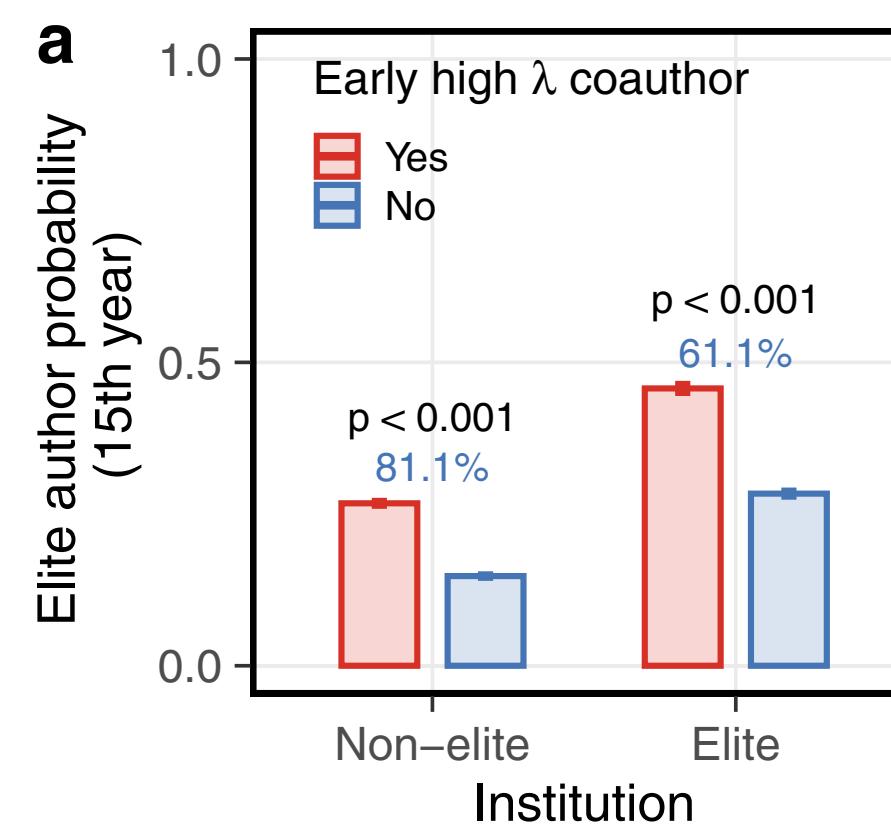
Pr. of high- θ early collab = 0.14 (elite) & 0.07 (non-elite)



elite institutions = top 10 by z-score of high impact papers
early-career collaboration = within first 5 years of publishing history
"elite author" = upper 5% of citations among authors in a given field-year
shaded areas are 95% confidence intervals

role of elite collaborators

- ▶ how does an early-career collaboration with an elite senior researcher influence prominence?
- elite senior researches with high λ or high θ



the 'benefits' are substantial regardless of coauthors career age

- ▶ slight decrease for most senior coauthors
- ▶ collaboration networks act like a partially transferrable form of *social capital* in science

✓

elite institutions = top 10 by z-score of high impact papers
early-career collaboration = within first 5 years of publishing history
"elite author" = upper 5% of citations among authors in a given field-year
shaded areas are 95% confidence intervals

how much of a meritocracy?

"little in academia makes sense except in the light of prestige"



- ▶ prestige pervades and structures the scientific ecosystem

how much of a meritocracy?

"little in academia makes sense except in the light of prestige"



- ▶ prestige pervades and structures the scientific ecosystem
 - shapes who gets faculty jobs and where
 - shapes how much scholarship is produced & academic labor
 - shapes composition of collaboration networks

how much of a meritocracy?

"little in academia makes sense except in the light of prestige"



▶ prestige pervades and structures the scientific ecosystem

- shapes who gets faculty jobs and where
- shapes how much scholarship is produced & academic labor
- shapes composition of collaboration networks

} individual placement, productivity, and prominence cannot be separated from one's place in the academic system

▶ your environment matters! (physical and social)

how much of a meritocracy?

"little in academia makes sense except in the light of prestige"



▶ prestige pervades and structures the scientific ecosystem

- shapes who gets faculty jobs and where
- shapes how much scholarship is produced & academic labor
- shapes composition of collaboration networks

} individual placement, productivity, and prominence cannot be separated from one's place in the academic system

▶ how can we accelerate science?

- deprioritize prestige signals
- deconcentrate academic labor
- targeted collaboration support, eg, for women & non-white scholars

▶ your environment matters!
(physical and social)

references & collaborators

Productivity, prominence, and the effects of academic environment

Samuel F. Way^{a,1}, Allison C. Morgan^a, Daniel B. Larremore^{a,b,2}, and Aaron Clauset^{a,b,c,1,2}

^a Department of Computer Science, University of Colorado, Boulder, CO, USA; ^b BioFrontiers Institute, University of Colorado, Boulder, CO, USA; ^c Santa Fe Institute, Santa Fe, NM, USA

PNAS 116(22), 10729–10733 (2019)

Quantifying hierarchy and dynamics in US faculty hiring and retention

<https://doi.org/10.1038/s41586-022-05222-x> K. Hunter Wapman¹, Sam Zhang², Aaron Clauset^{1,3,4} & Daniel B. Larremore^{1,3}

Nature 610, 120–127 (2022)

Labor advantages drive the greater productivity of faculty at elite universities

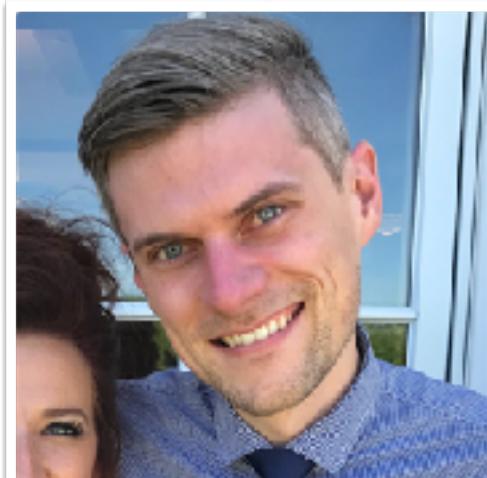
Sam Zhang^{1*}, K. Hunter Wapman², Daniel B. Larremore^{2,3}, Aaron Clauset^{2,3,4*}

Science Advances 8, eabq7056 (2022)

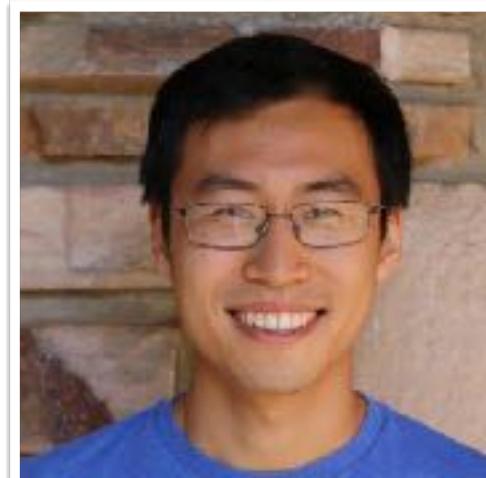
Untangling the network effects of productivity and prominence among scientists

Weihua Li^{1,2,3,4}, Sam Zhang⁵, Zhiming Zheng^{1,2,3,4}, Skyler J. Cranmer⁶ & Aaron Clauset^{1,7,8,9}

Nature Communications 13, 4907 (2022)



Dr. Samuel F Way
(now: Spotify)



Sam Zhang
(Colorado)



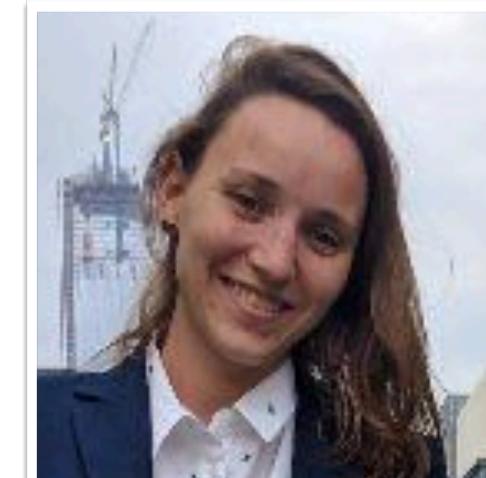
K. Hunter Wapman
(Colorado)



Prof. Weihua Li
(Beihang)



Prof. Zhiming Zhang
(Beihang)



Dr. Allison Morgan
(now: Code for America)



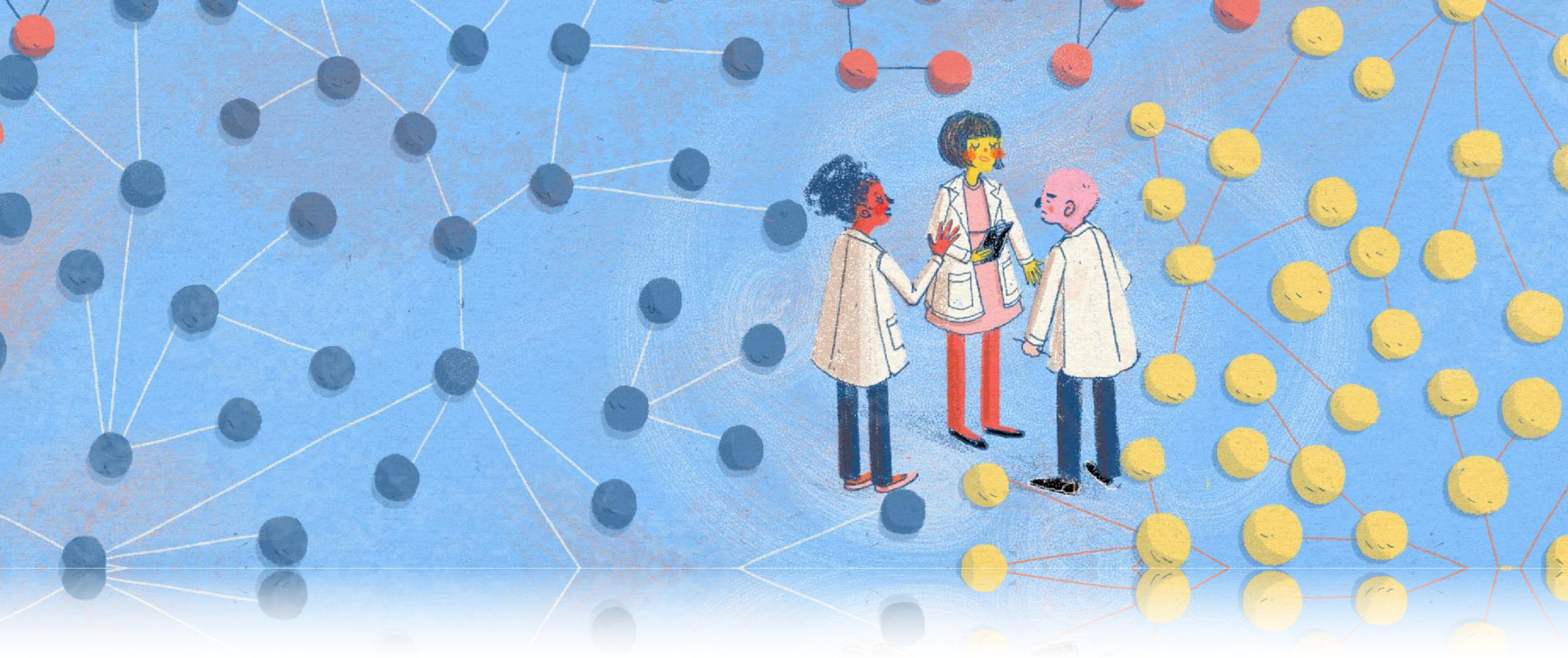
Prof. Skyler Cranmer
(Ohio State)



Prof. Daniel Larremore
(Colorado)

Funding:





fin



papers, code, data

<https://aaronclauset.github.io>

what is prestige?

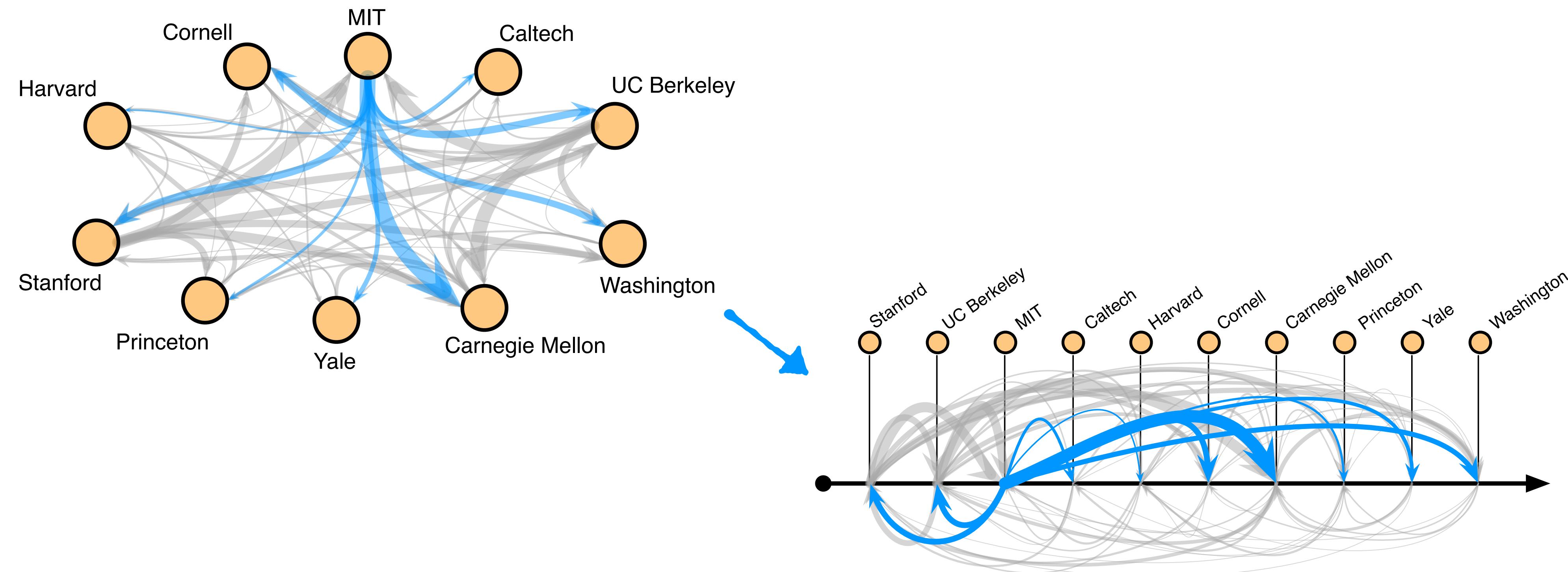
prestige in the sense of Burris (2004) : a form of social capital

which we can infer from *who hires whose graduates as faculty*

prestige & the scientific workforce

faculty hiring is a *network*

prestige → centrality in the faculty hiring network ("placement power")



prestige & the scientific workforce

faculty hiring is a *network*

prestige → centrality in the faculty hiring network ("placement power")



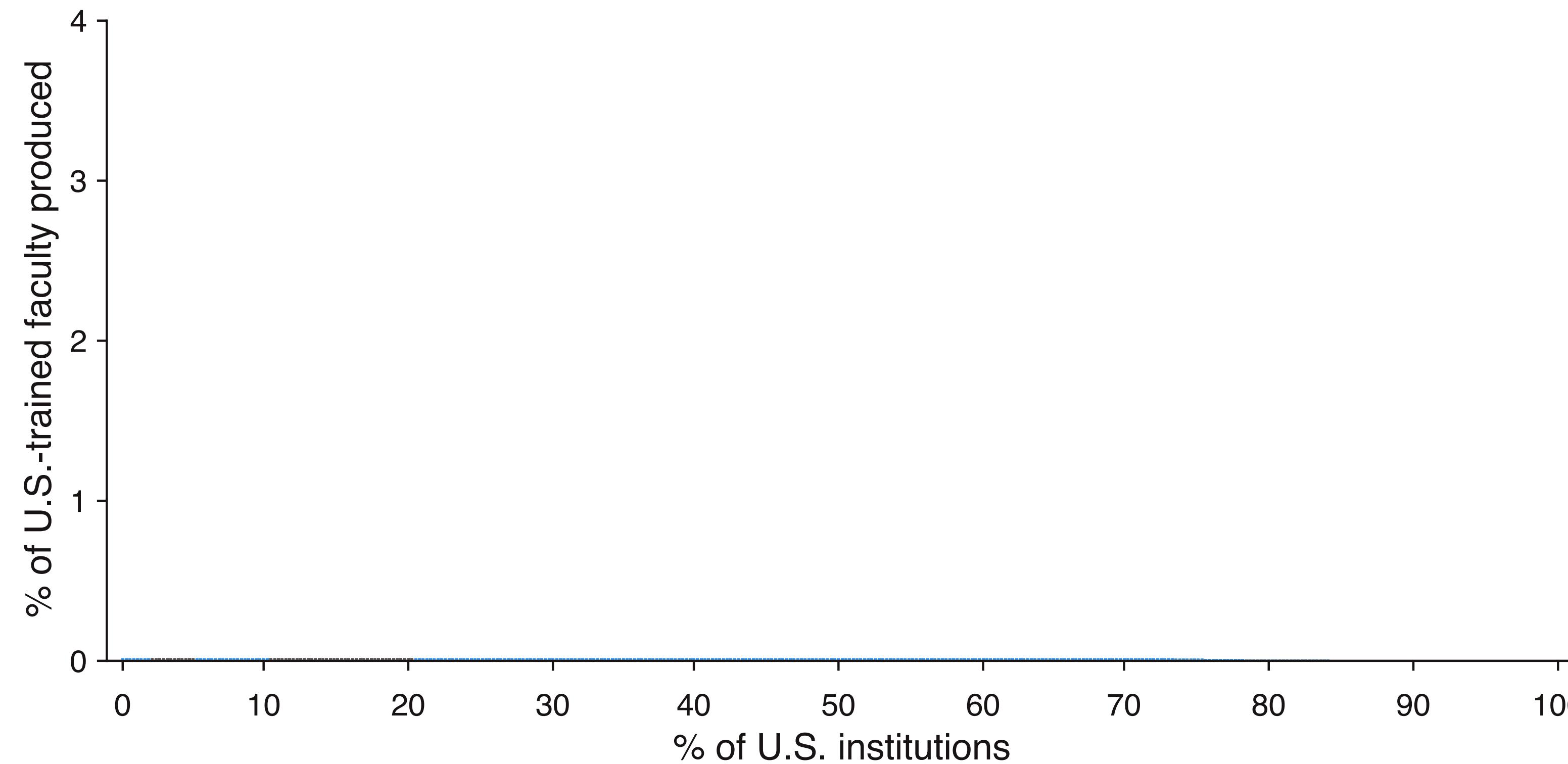
census-level data on education and placement of 291,123 regular faculty at 10,612 departments across 86 fields, 2011 – 2020



- ▶ how unequal is faculty production?
- ▶ what implications for epistemic inequalities?

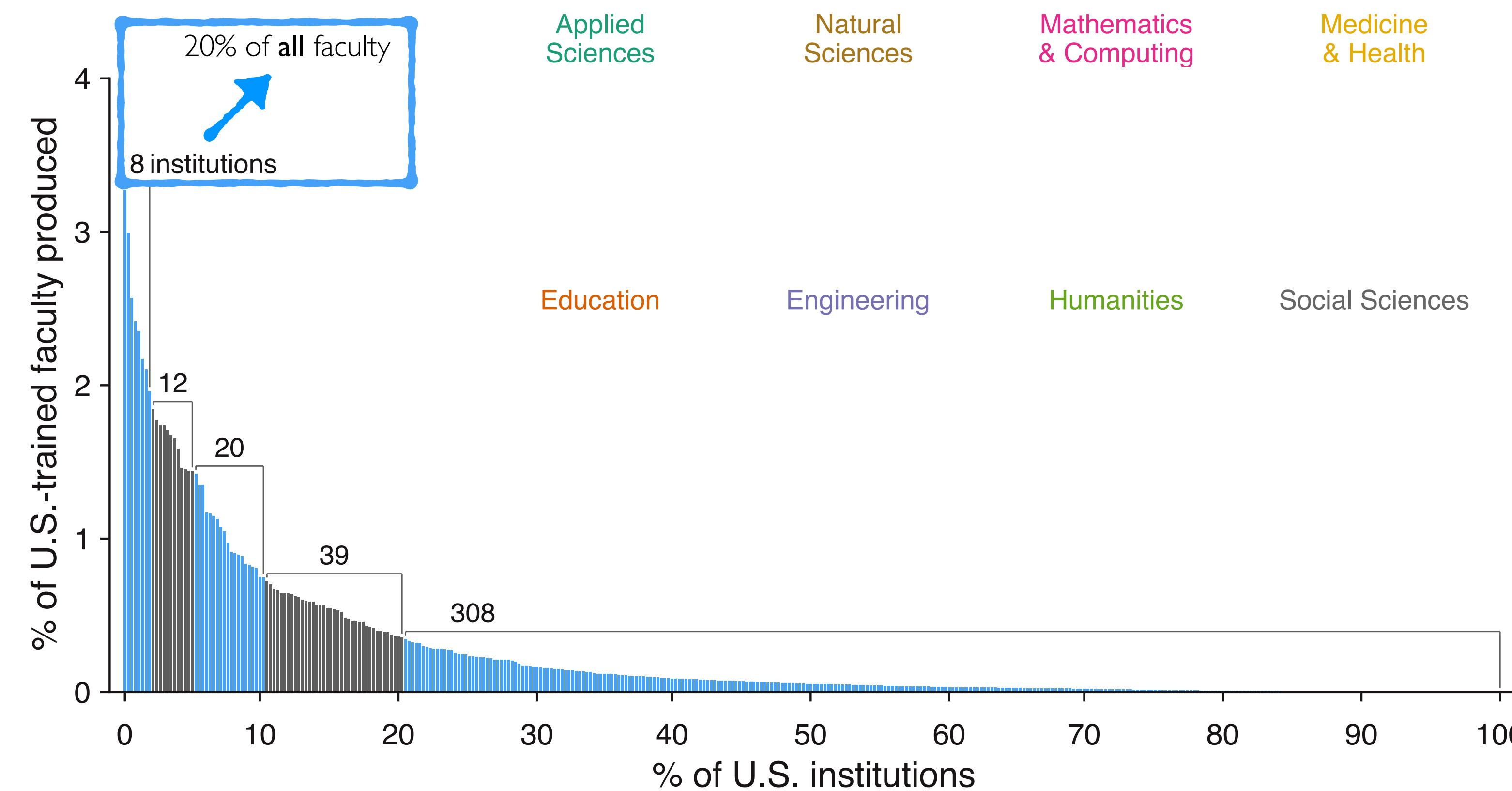
who hires whose graduates as faculty?

- ▶ sort 387 institutions by overall production of faculty



who hires whose graduates as faculty?

- faculty production is enormously *concentrated*



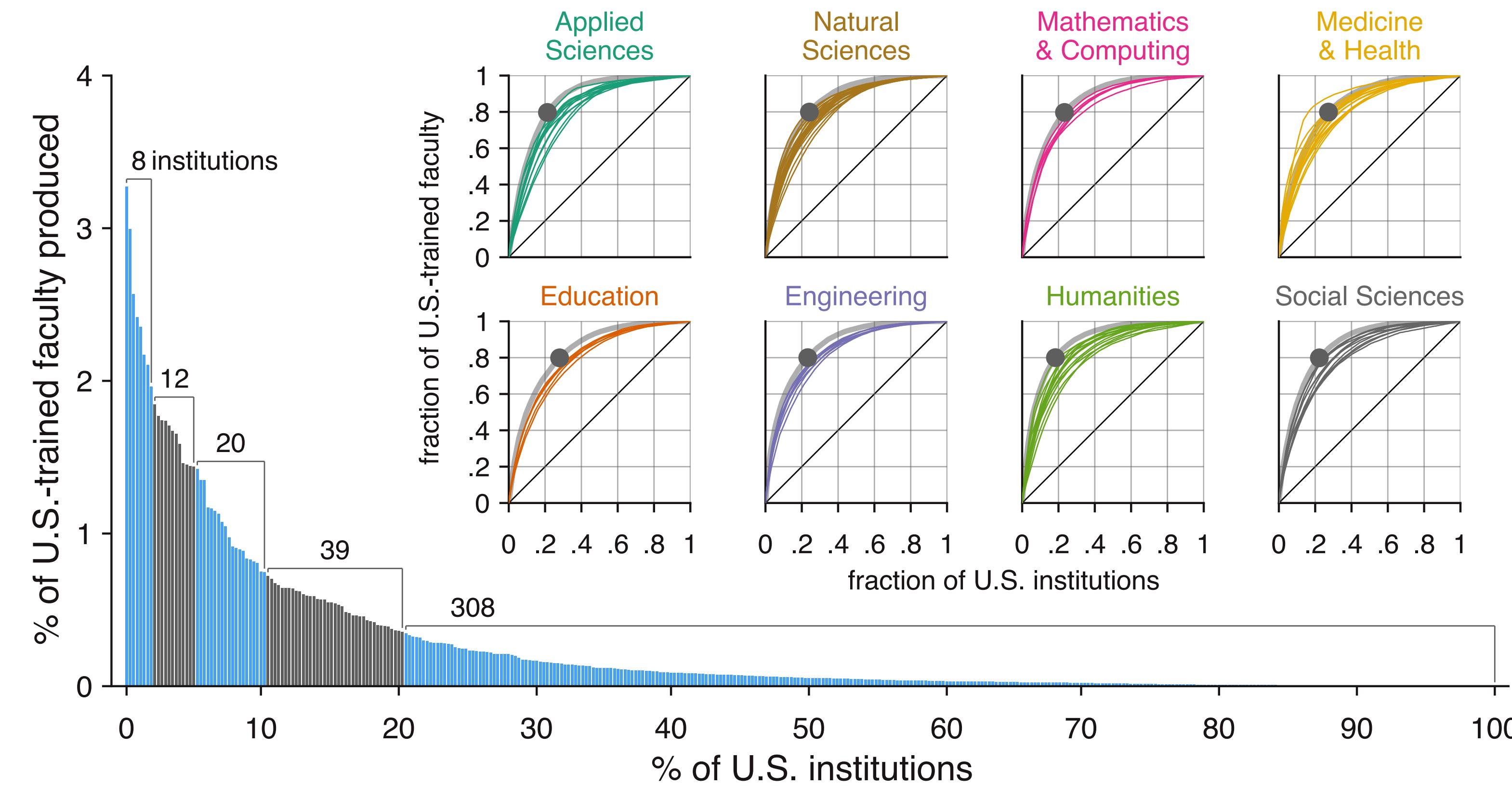
all-academia Gini = 0.75

the top 8: Berkeley, Harvard, Michigan, Wisconsin, Stanford, UIUC, MIT, UT Austin (note that only 3 of the 8 are private)

Wapman et al., "Quantifying hierarchy and dynamics in US faculty hiring and retention" (2022)

who hires whose graduates as faculty?

- ▶ faculty production is enormously *concentrated*
- ▶ recapitulated in all 86 fields (a roughly universal "80-20 rule")



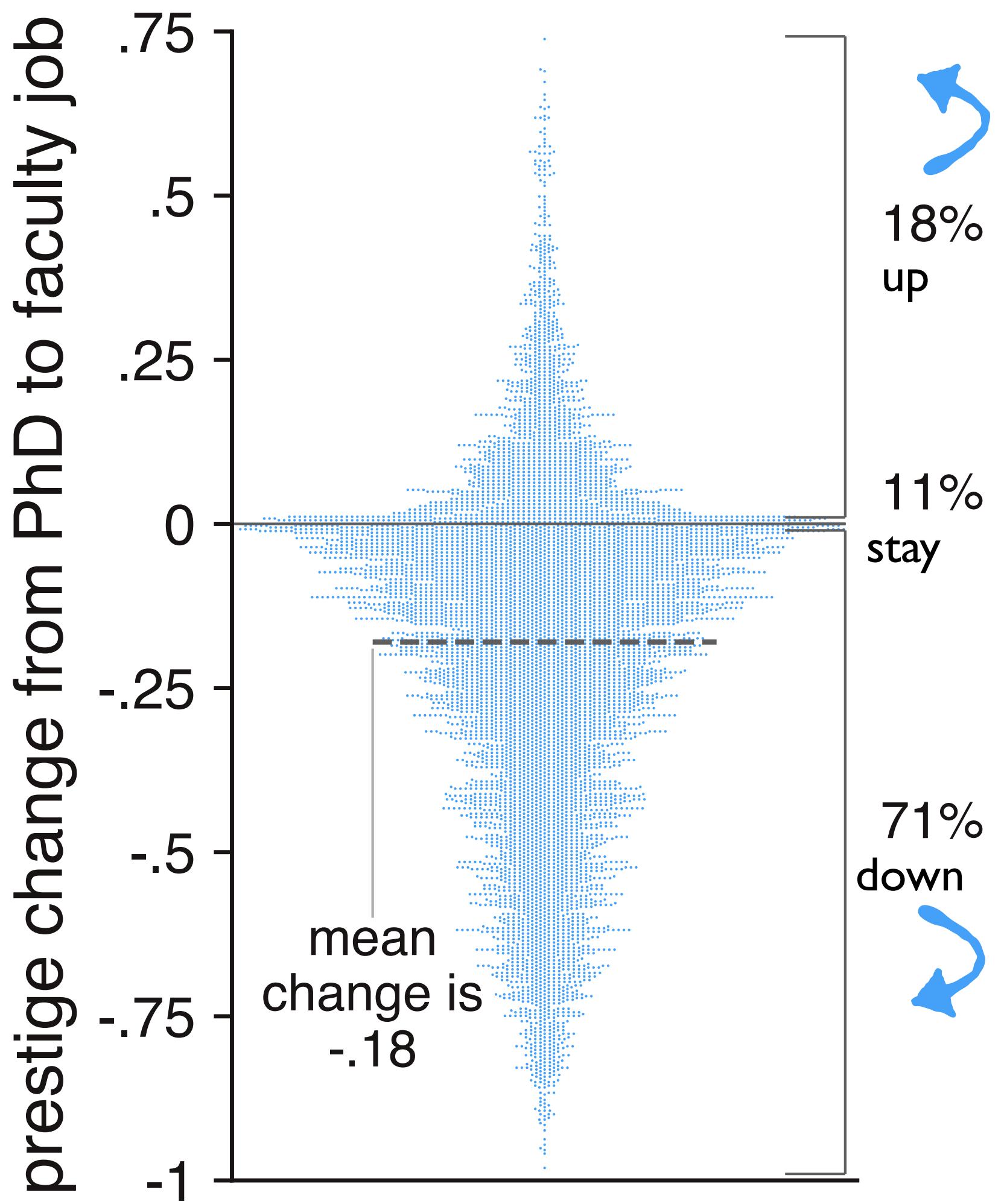
all-academia Gini = 0.75

the top 8: Berkeley, Harvard, Michigan, Wisconsin, Stanford, UIUC, MIT, UT Austin (note that only 3 of the 8 are private)

Wapman et al., "Quantifying hierarchy and dynamics in US faculty hiring and retention" (2022)

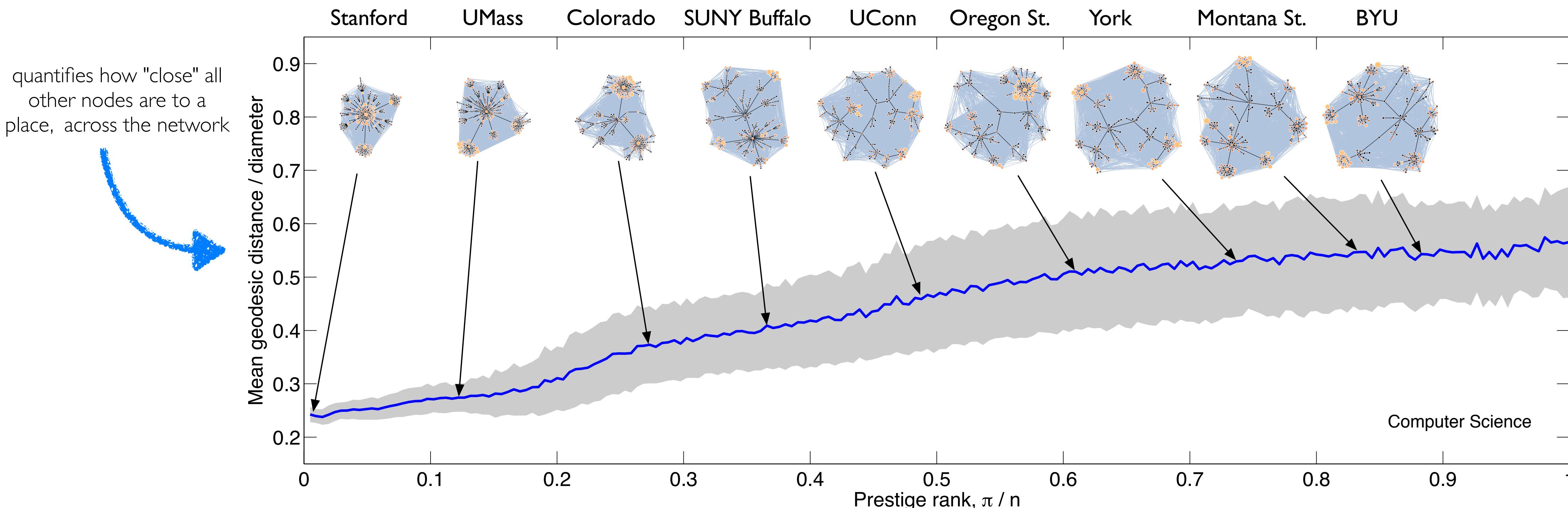
who hires whose graduates as faculty?

- ▶ prestige hierarchies are steep
- ▶ faculty placement mostly "down" (71%)
only 18% (mean) move "up"
- ▶ only 20% of departments have trained
more faculty than they've hired



who hires whose graduates as faculty?

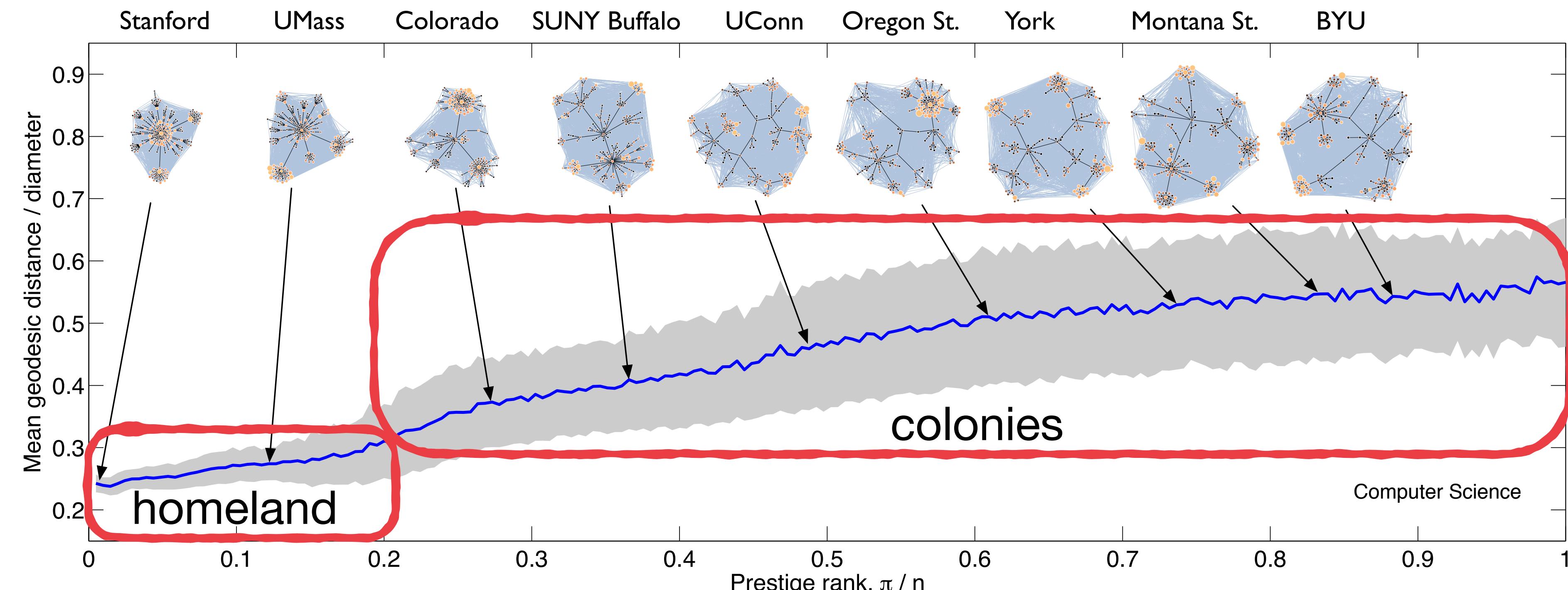
network organized as core and periphery : high ranked nodes are more central



who hires whose graduates as faculty?

network organized as ~~core and periphery~~ *homeland* and *colonies*

prestige → *influence* via doctoral placement
over research agendas, research communities, and departmental norms across a field



who hires whose graduates as faculty?

- ▶ prestige is a *structural variable* in the science of science
- placement power π quantifies reputation via outcomes (not inputs)
- reveals *core-periphery* structure of academia
 - faculty flow from core → periphery ("the colonies")
 - modest fraction stays inside core ("homeland")
 - small fraction flows "upstream"
 - these hierarchies extremely stable over time
- prestige → faculty production → hierarchy

