



# **Networks, Prestige, and the Spread of Scientific Ideas**

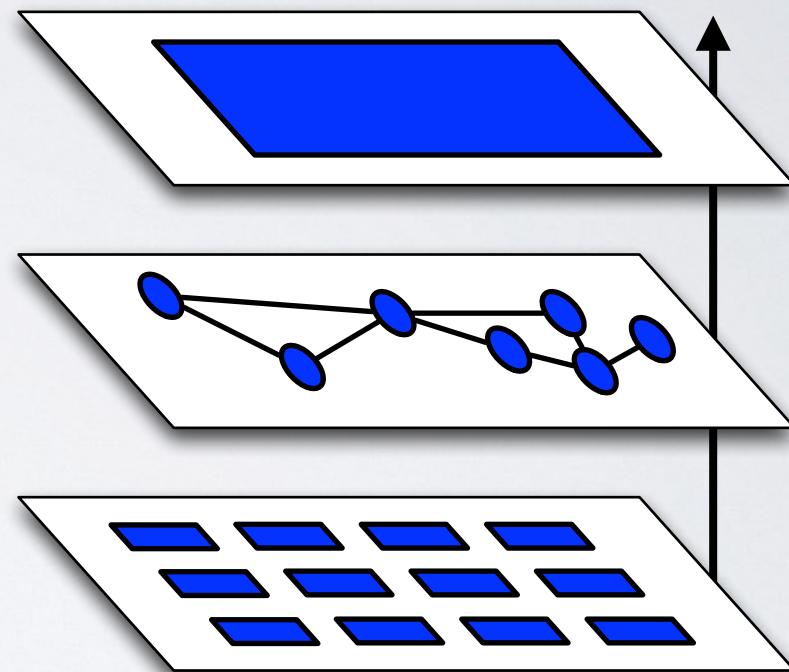
Aaron Clauset  
@aaronclauset  
Computer Science Dept. & BioFrontiers Institute  
University of Colorado, Boulder  
External Faculty, Santa Fe Institute

**what are networks?**

# **what are networks?**

- an approach
- representation of structural complexity
- connect "micro" to "macro"
- *structure above*  
individuals / components
- *structure below*  
system / population

## **system / population**



## **individuals / components**

## **learning goals**

- build intuition
- expose key concepts
- highlight some big questions
- use "science" as a model complex system
- apply network ideas
- not a substitute for technical coursework

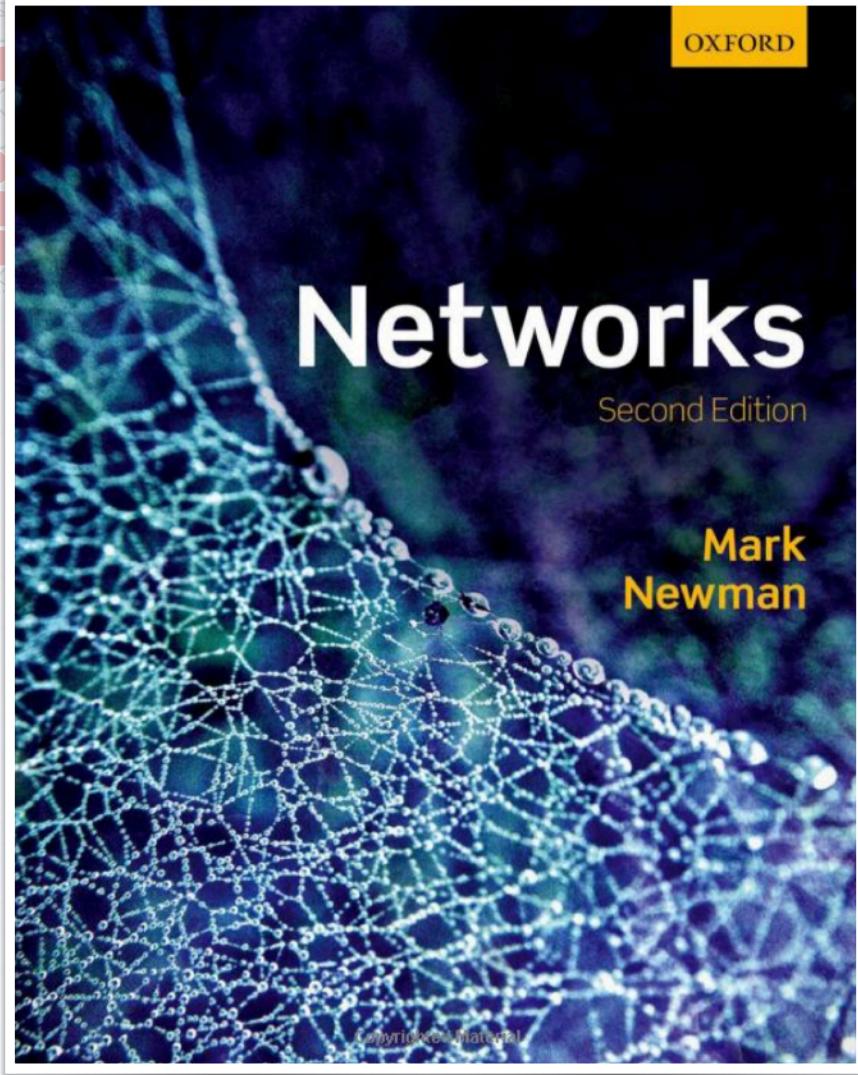


Mark Newman

Professor of Physics  
University of Michigan

External Faculty  
Santa Fe Institute

<http://www-personal.umich.edu/~mejn/>





University of Colorado **Boulder**

## **Network Analysis and Modeling**

Instructor: Aaron Clauset or Daniel B. Larremore

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/>



University of Colorado **Boulder**

## **Biological Networks**

Instructor: Aaron Clauset

This undergraduate-level course examines the computational representation and analysis of biological phenomena through the structure and dynamics of networks, from molecules to species. Attention focuses on algorithms for clustering network structures, predicting missing information, modeling flows, regulation, and spreading-process dynamics, examining the evolution of network structure, and developing intuition for how network structure and dynamics relate to biological phenomena.

*Full lectures notes online (~150 pages in PDF)*

<https://aaronclauset.github.io/courses/3352/>

## Software

R

Python

Matlab



NetworkX [python]

graph-tool [python, c++]

GraphLab [python, c++]

## Standalone editors

UCI-Net

NodeXL

Gephi

Pajek

Network Workbench

Cytoscape

yEd graph editor

Graphviz

## Network data sets



Colorado Index of Complex Networks

The screenshot shows a web browser window displaying the 'Index of Complex Networks' website. The address bar shows 'icon.colorado.edu/#!/'. The page has a dark header with the title 'Index of Complex Networks' and navigation links for 'NETWORKS', 'ABOUT', and 'SUGGEST...'.

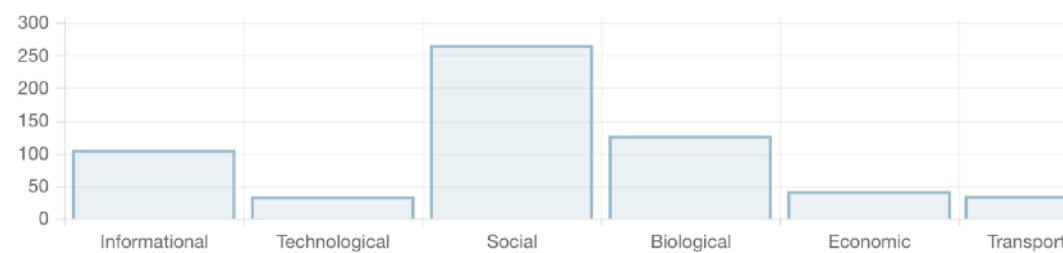
### The Colorado Index of Complex Networks (ICON)

ICON is a comprehensive index of research-quality network data sets from all domains of networks, 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 on the [NETWORKS tab](#) above to get started.

Entries found: 609 Networks found: 4419



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

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

**prestige pervades and structures the scientific ecosystem**

- shapes who gets jobs and where
- shapes who publishes where
- shapes who works on what ideas
- shapes who gets what resources (grants, students, attention, etc.)
- etc.

 how many good ideas are left unexplored because  
they lack the right markers of prestige?

## what is prestige?

- prestige is related to *rankings*, but
- everyone agrees that USNews and NRC rankings are bad [these measure *inputs* to science, not *outputs* or *value-added*] 

## what is prestige?

- prestige is related to *rankings*, but
- everyone agrees that USNews and NRC rankings are bad 
- we can extract a *data-driven ranking* from the network of who hires whose graduates as faculty 

vertices are PhD-granting universities

consumers  $\leftrightarrow$  producers

$v$  hires from  $u$ , add an edge  $u \rightarrow v$

- prestige  $\rightarrow$  placement power  $\rightarrow$  the ability to place your graduates as faculty at other universities
- to measure prestige, we need faculty hiring data

**COLLECT**



**ALL THE DATA**

memegenerator.net

## collect all the data (2015)

complete, hand-curated data  
for 19,000 tenure-track faculty  
across 461 departments in

- Computer Science (205 depts)
- Business (112)
- History (144)

roughly 4000 hours of manual  
data collection

```
>>> record 1059
# facultyName : James H. Martin
# email      :
# sex        : M
# department : Computer Science
# place      : University of Colorado, Boulder
# current    : Full Professor
# [Education]
# degree     : BS
# place      : Columbia University
# field      : Computer Science
# years      : ????-1981
# [Education]
# degree     : PhD
# place      : UC Berkeley
# field      : Computer Science
# years      : ????-1988
# [Faculty]
# rank       : Assistant Professor
# place      : University of Colorado, Boulder
# years      : 1989-1995
# [Faculty]
# rank       : Associate Professor
# place      : University of Colorado, Boulder
# years      : 1995-2007
# [Faculty]
# rank       : Full Professor
# place      : University of Colorado, Boulder
# years      : 2007-2011
# recordDate : 7/4/2011
```

## collect all the data (2015)

complete, hand-curated data  
for 19,000 tenure-track faculty  
across 461 departments in

- Computer Science (205 depts)
- Business (112)
- History (144)

roughly 4000 hours of manual  
data collection

```
>>> record 1059
# facultyName : James H. Martin
# email      :
# sex        : M
# department : Computer Science
# place      : University of Colorado, Boulder
# current    : Full Professor
# [Education]
# degree     : BS
# place      : Columbia University
# field      : Computer Science
# years      : ????-1981

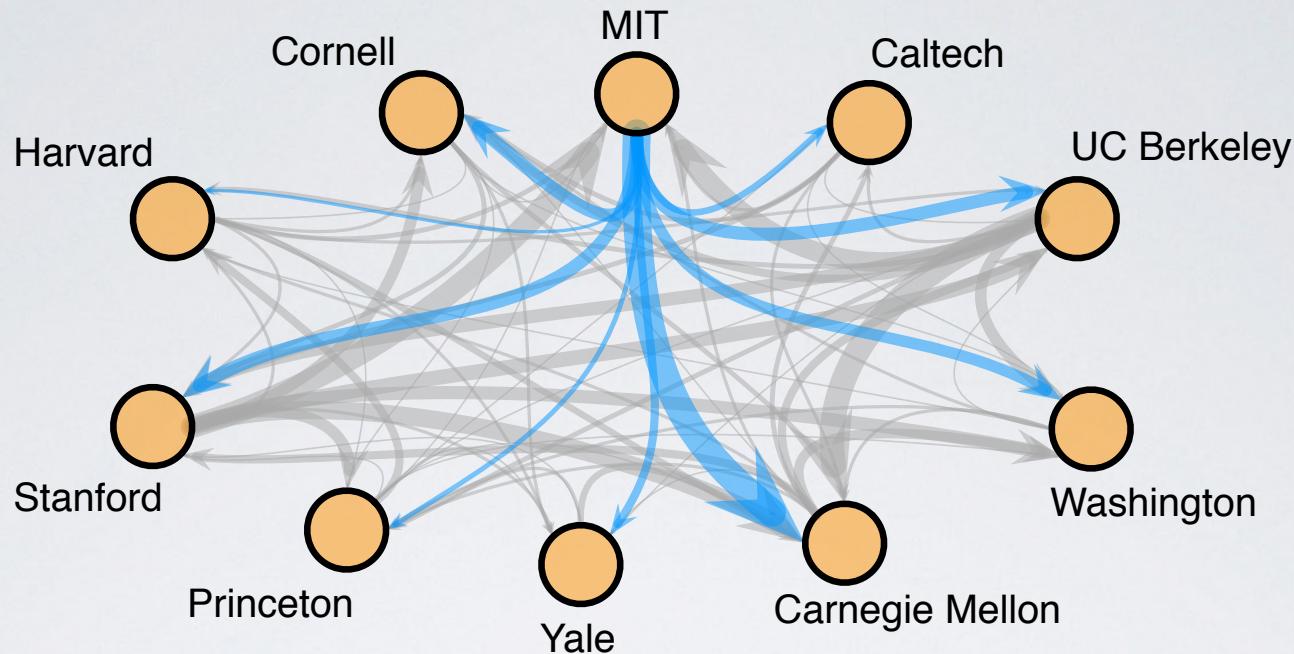
# [Education]
# degree     : PhD
# place      : UC Berkeley
# field      : Computer Science
# years      : ????-1988

# [Faculty]
# rank       : Assistant Professor
# place      : University of Colorado, Boulder
# years      : 1989-1995

# [Faculty]
# rank       : Associate Professor
# place      : University of Colorado, Boulder
# years      : 1995-2007

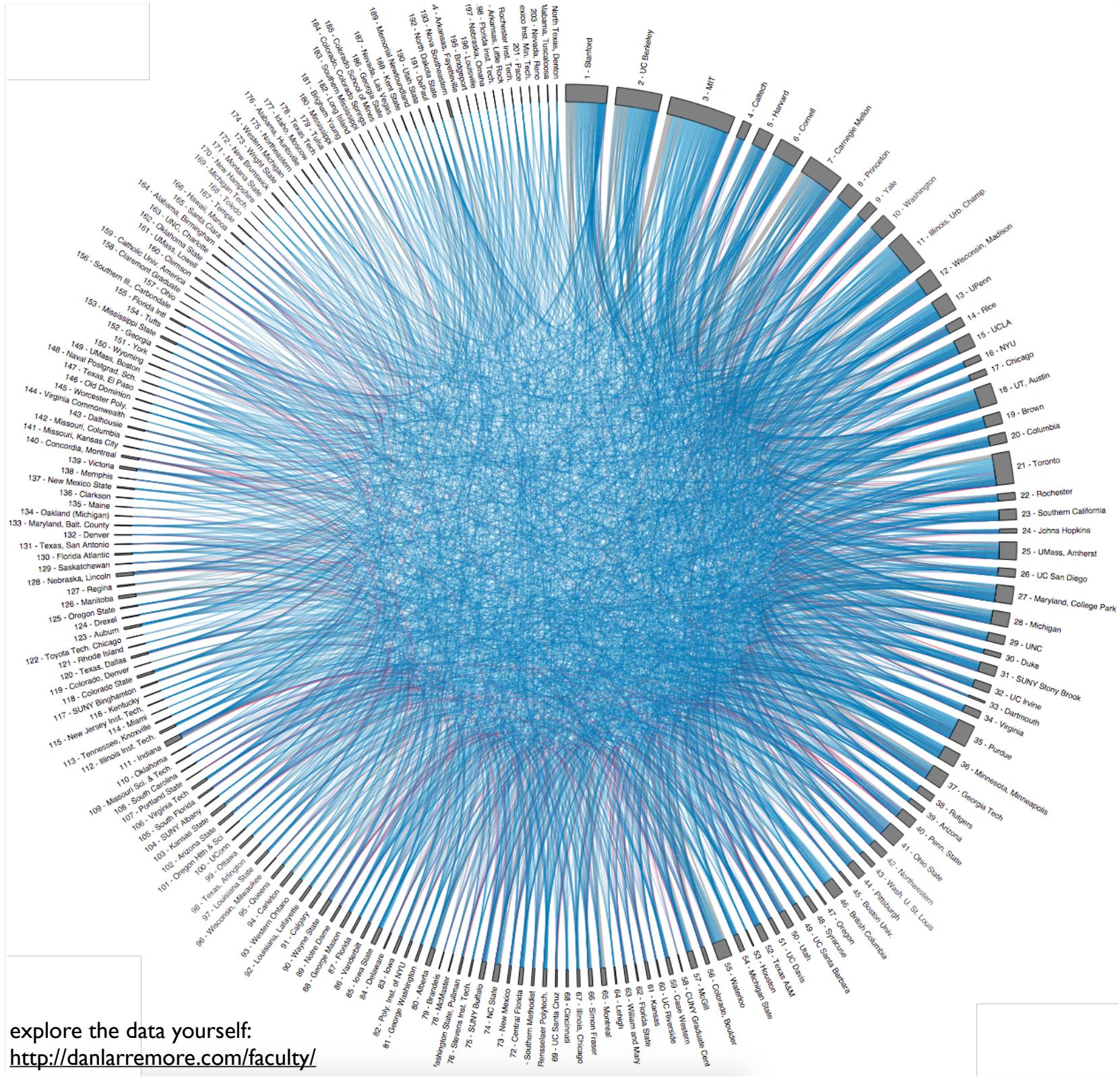
# [Faculty]
# rank       : Full Professor
# place      : University of Colorado, Boulder
# years      : 2007-2011
# recordDate : 7/4/2011
```





## faculty hiring is a *network*

- vertices are PhD-granting universities
- consumers  $\leftrightarrow$  producers
- $v$  hires from  $u$ , add an edge  $u \rightarrow v$



explore the data yourself:  
<http://danlarremore.com/faculty/>

**huge inequalities in faculty production**

## huge inequalities in faculty production

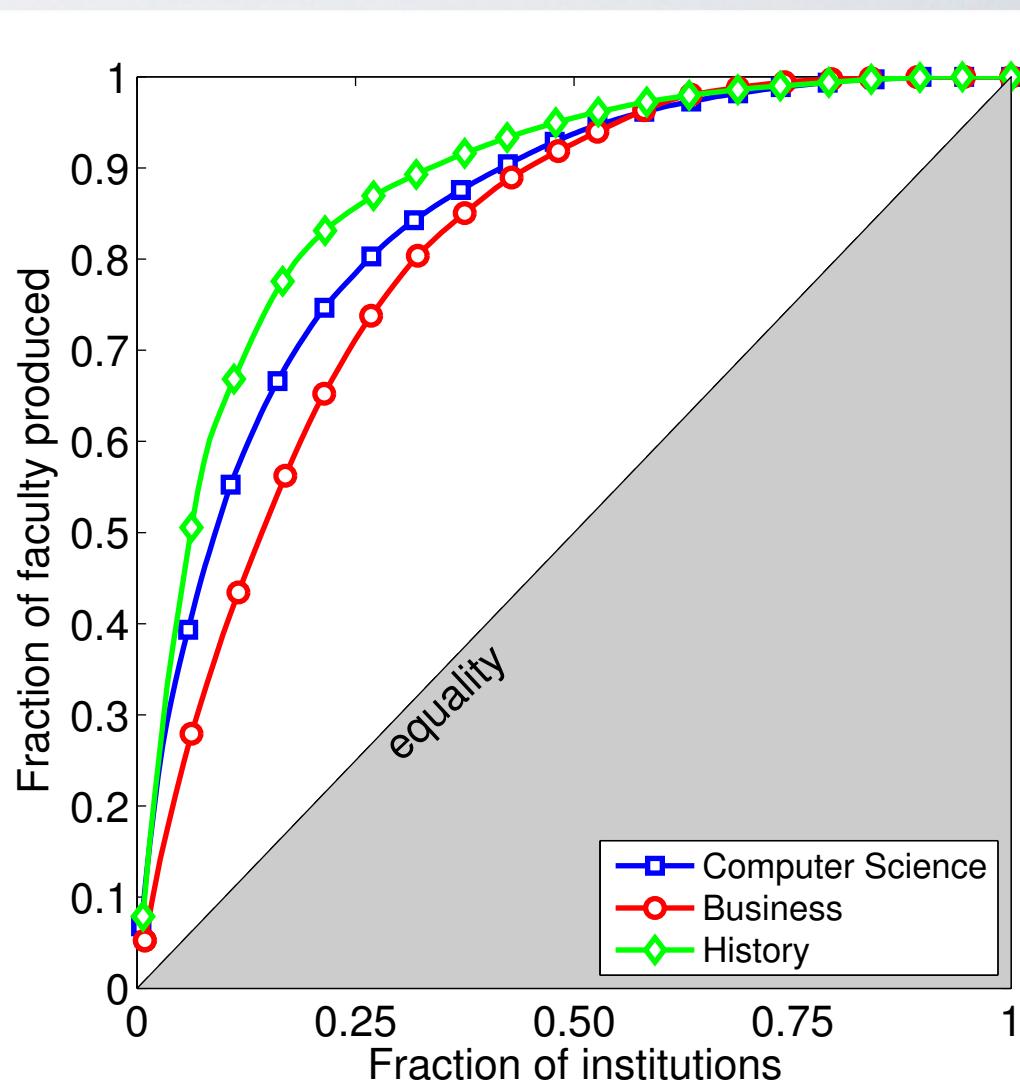
### Computer Science (205 depts.)

faculty production, Gini = 0.69

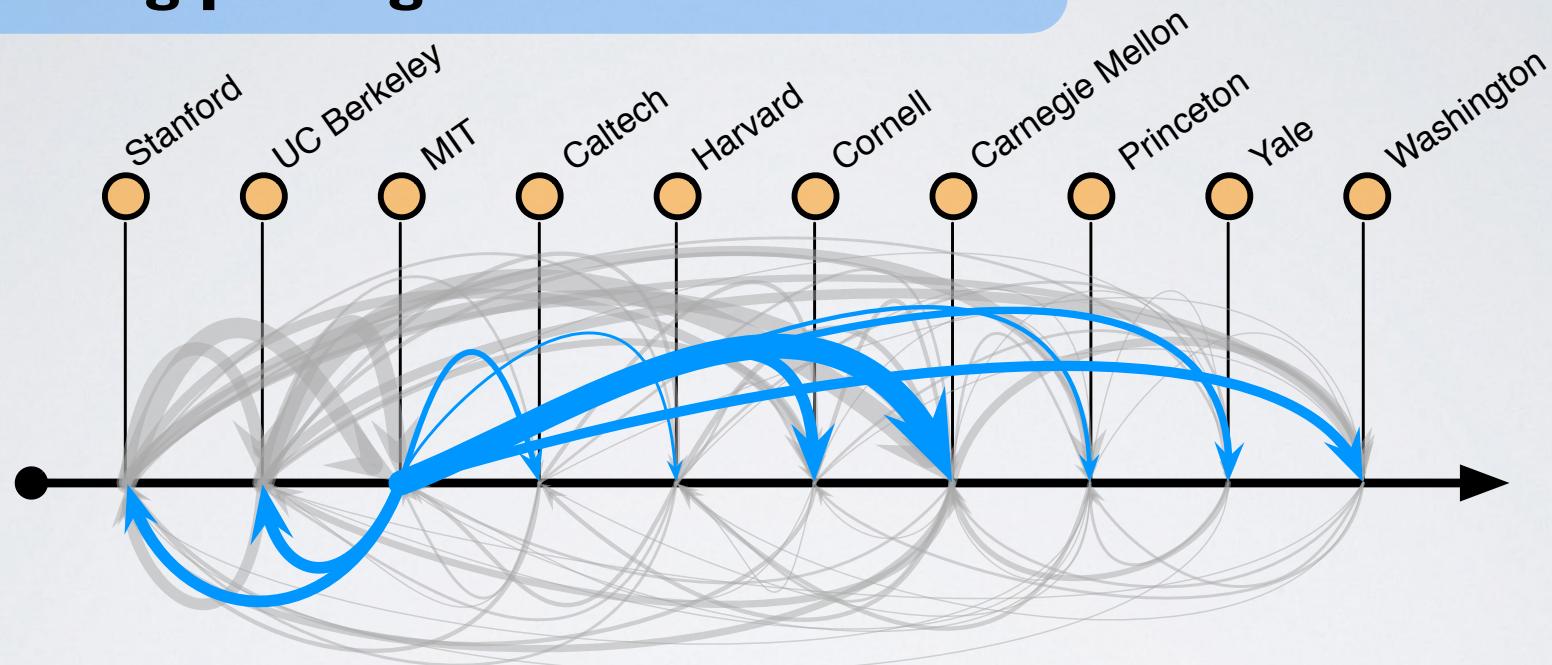
18 depts. (9%) produced 50% of *all* faculty

76% of depts. = net consumers  
[more faculty hired than produced]

depts. ranked 1-10 produce :  
1.6x more than ranks 11-20  
3.1x more than ranks 21-30

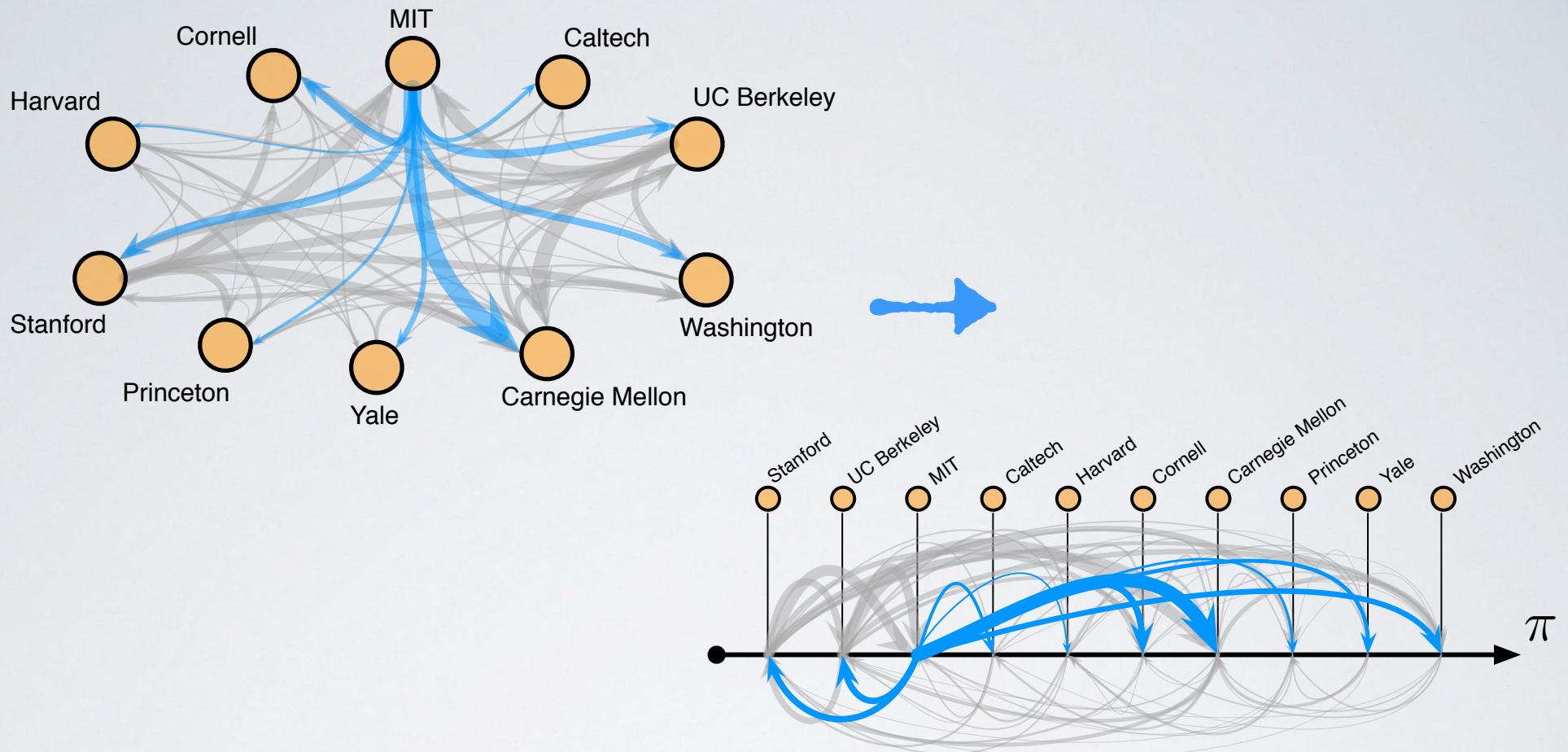


## measuring prestige



- select permutation (a ranking)  $\pi$  that minimizes the number of "rank violations" : edges  $(u, v)$  where  $\pi_v < \pi_u$
- higher-ranked nodes have greater *placement power*
- Minimum Violation Rankings (MVRs) are equivalent to solving *minimum feedback arc set problem* (NP-hard)

these "MVR"s have a deep history in social theory for extracting dominance or prestige hierarchies from data, especially in animal behavior (see [de Vries \(1999\)](#))  
there are many equivalent MVRs for our network, we sample these using a zero-temp MCMC, and average across them to obtain  $\langle \pi \rangle$



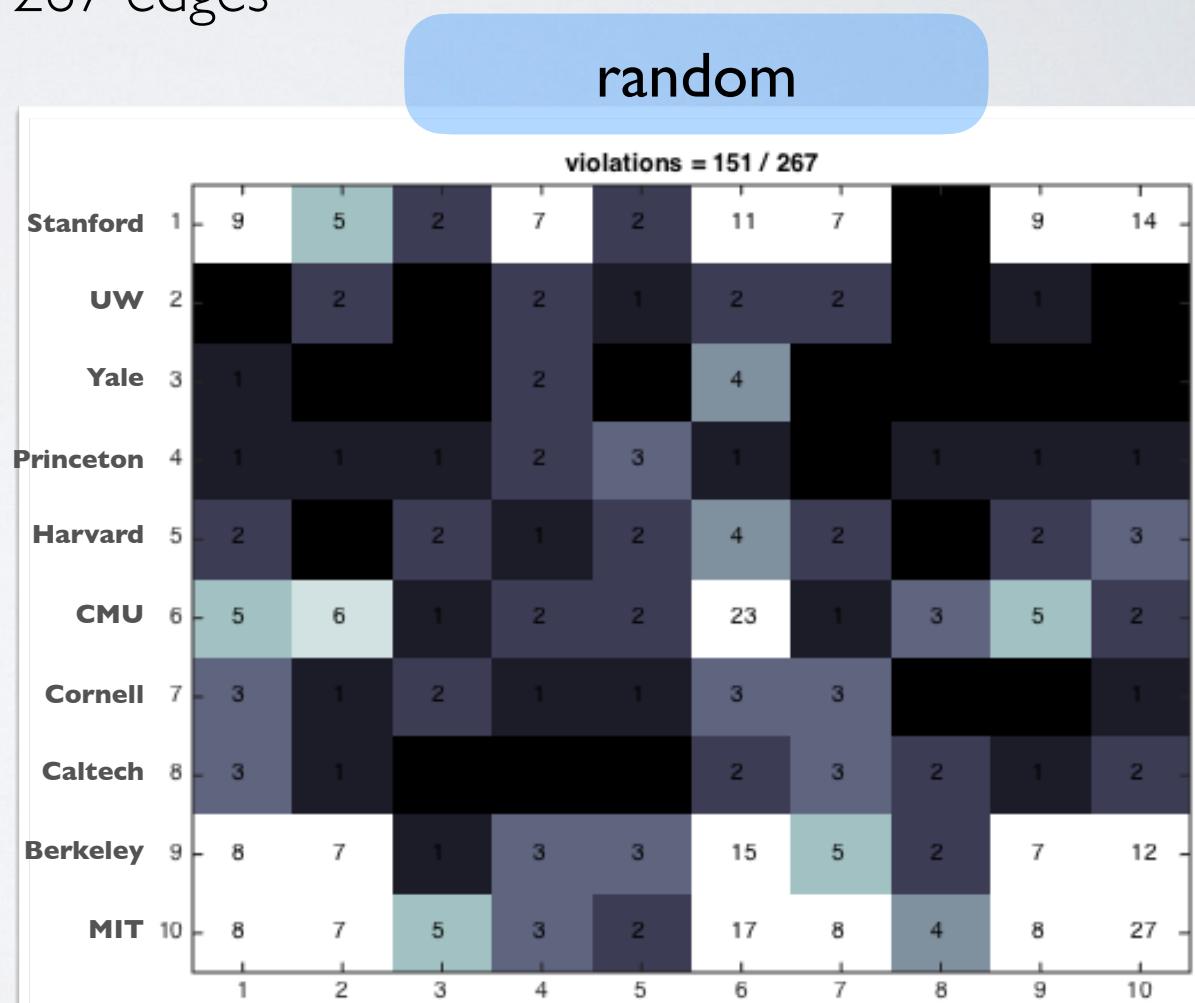
- the ordering  $\pi$  of the nodes that minimizes rank violations over arcs:

$$\text{MVR}(A) = \inf_{\pi} \sum_{u,v} A_{u,v} \times \text{sign}(\pi_u - \pi_v)$$

- actually, many such  $\pi$ 's, so we sample them using MCMC and compute the mean ranking  $\langle \pi \rangle$

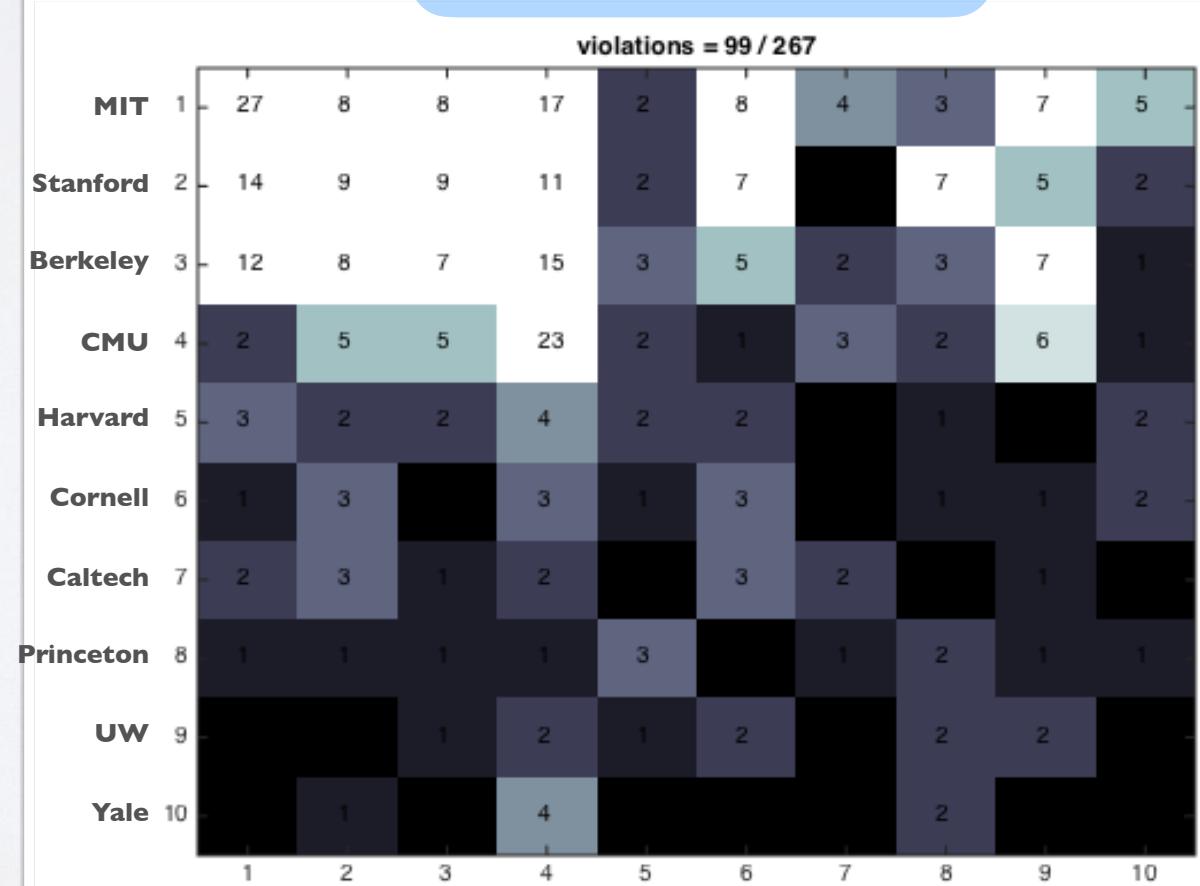
these "MVR"s have a deep history in social theory for extracting dominance or prestige hierarchies from data, especially in animal behavior (see [de Vries \(1999\)](#))  
 there are many equivalent MVRs for our network. we sample these using a zero-temp MCMC, and average across them to obtain

- given an ordering  $\pi$  with  $\psi(\pi, A)$  rank violations on network  $A$
- for instance: 151 "violations" of 267 edges

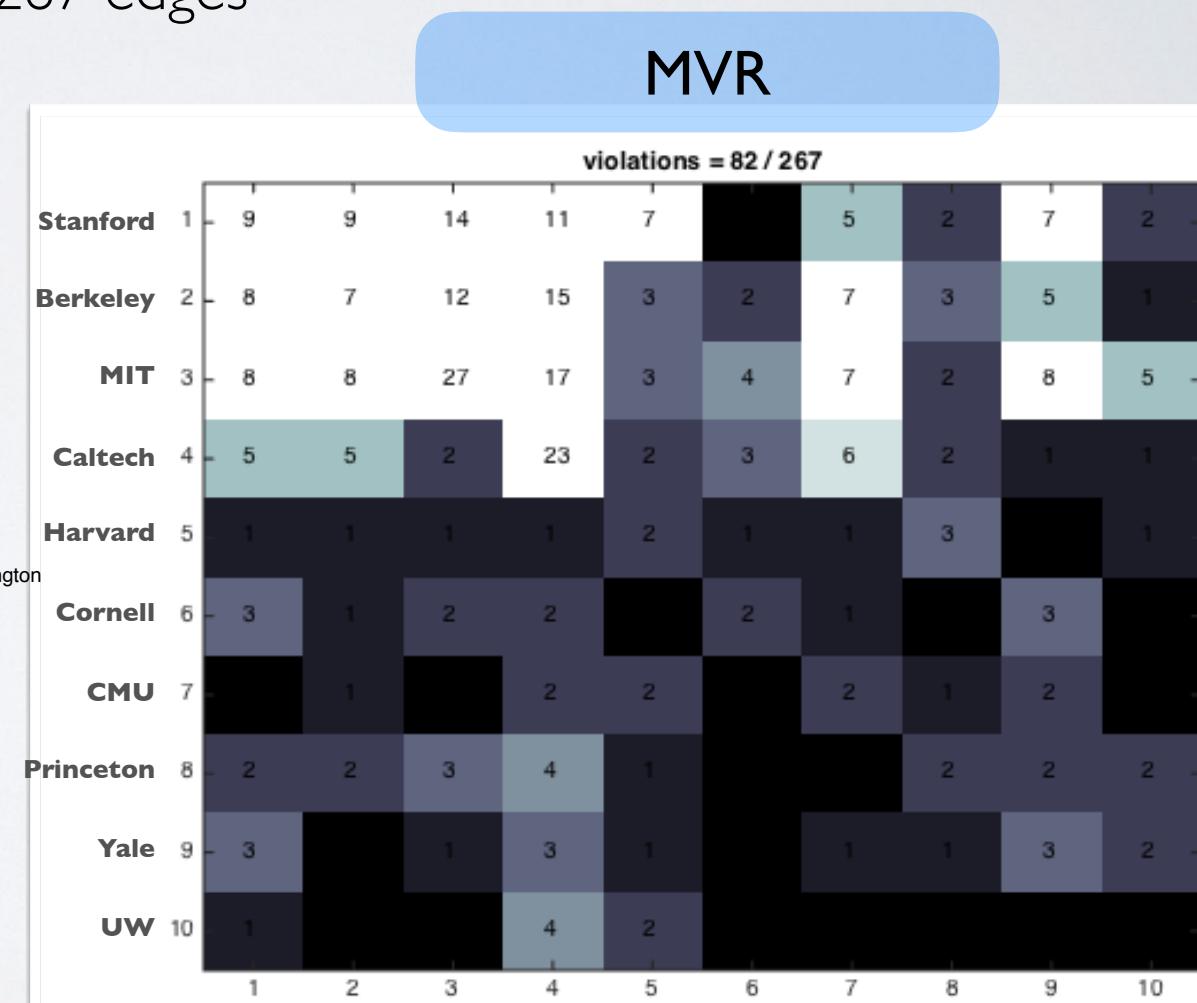
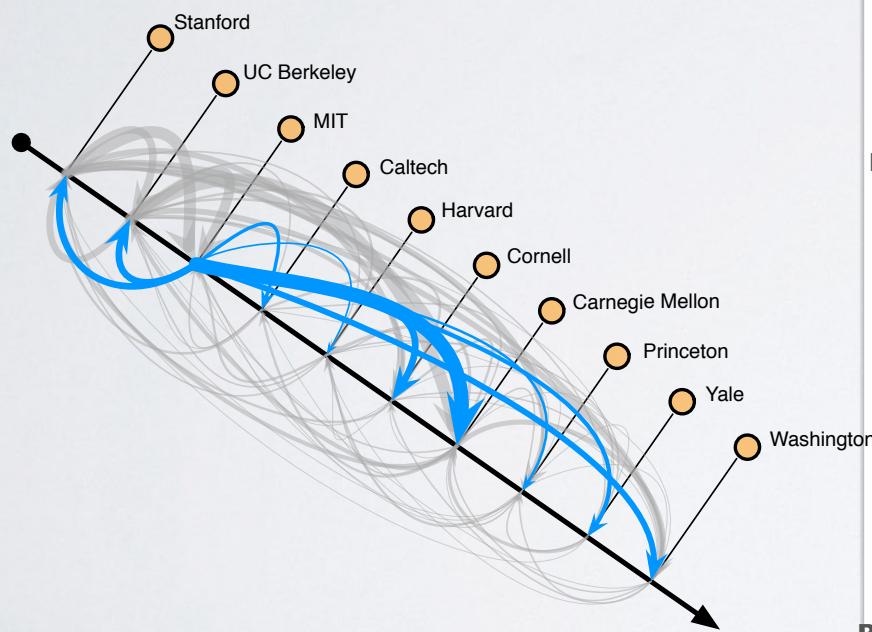


- given an ordering  $\pi$  with  $\psi(\pi, A)$  rank violations on network  $A$
- for instance: 99 "violations" of 267 edges

sort by out-degree

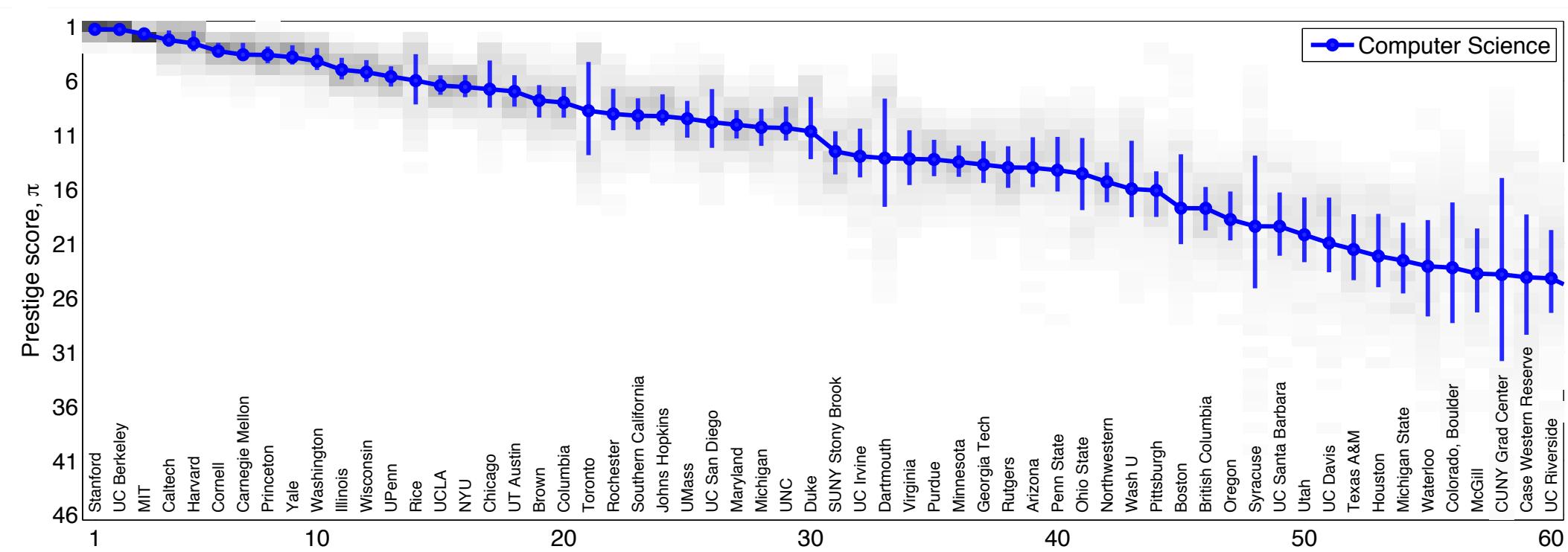


- given an ordering  $\pi$  with  $\psi(\pi, A)$  rank violations on network  $A$
- MCMC sampler : choose a pair  $(u, v)$ , swap their ranks  $\pi_u \leftrightarrow \pi_v$  to obtain  $\pi'$ , compute  $\psi(\pi', A)$ , accept change if  $\psi(\pi', A) \geq \psi(\pi, A)$
- for instance: 82 "violations" of 267 edges



## **measuring prestige**

- what do these "prestige" hierarchies look like?
- what do they tell us about the structure of the faculty market?



prestige correlates with USNews / NRC, but is *more predictive* of placement

Computer Science	$\pi$	US News	NRC	$\lambda$
$\pi$	—	0.79 (153)	0.80 (103)	0.80 (205)
US News	0.79 (153)	—	0.83 (103)	0.74 (153)
NRC	0.80 (103)	0.83 (103)	—	0.81 (103)
$\lambda$	0.80 (205)	0.74 (153)	0.81 (103)	—

- prestige  $\pi$  directly measures *placement power*
- uncertainty increases as prestige decreases
- similar results, but different orderings for *Business* and *History*

## most placements far down the hierarchy

### Computer Science

80% move down

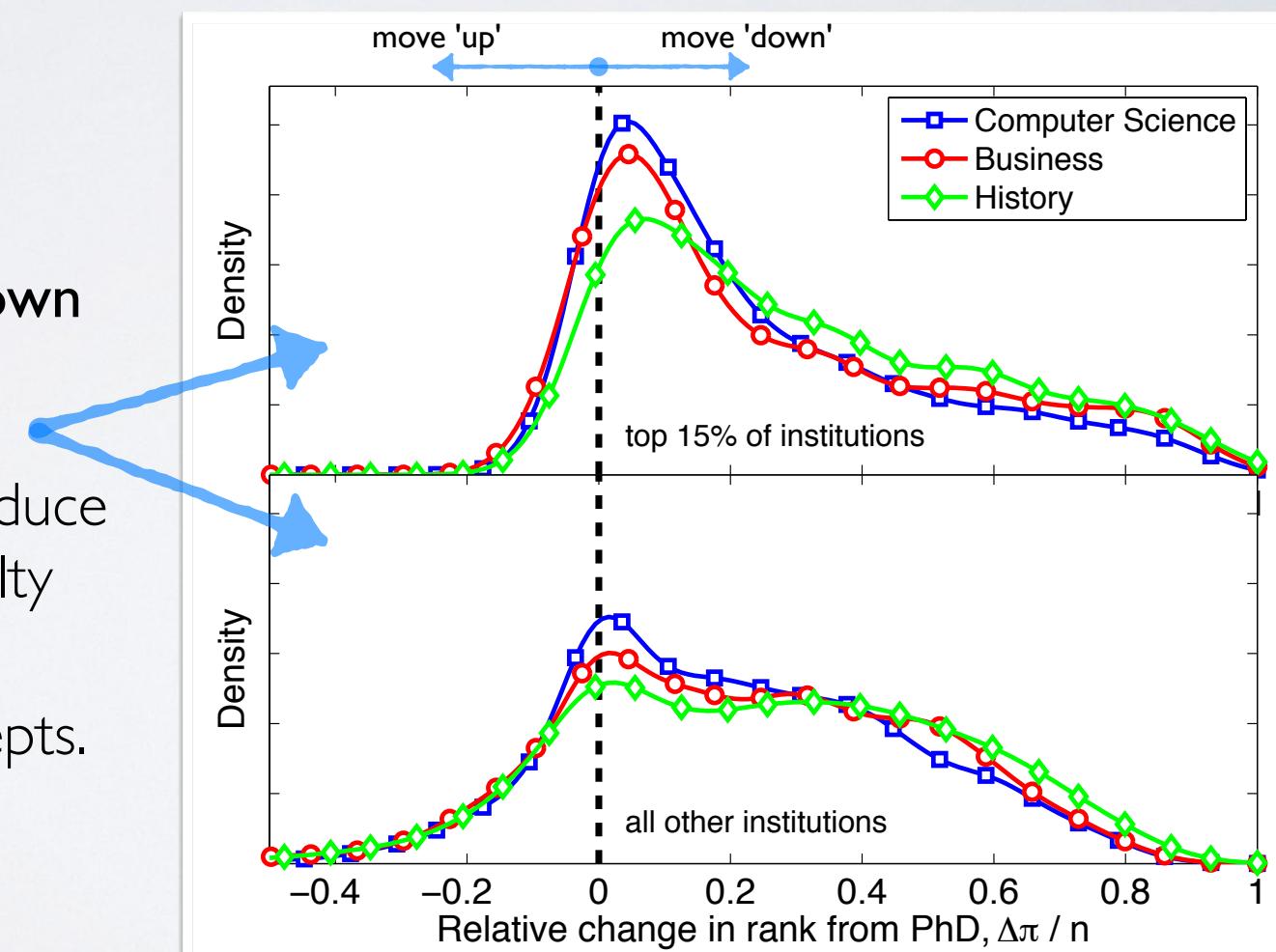
8% self-hires

12% move up

average = 47 steps down

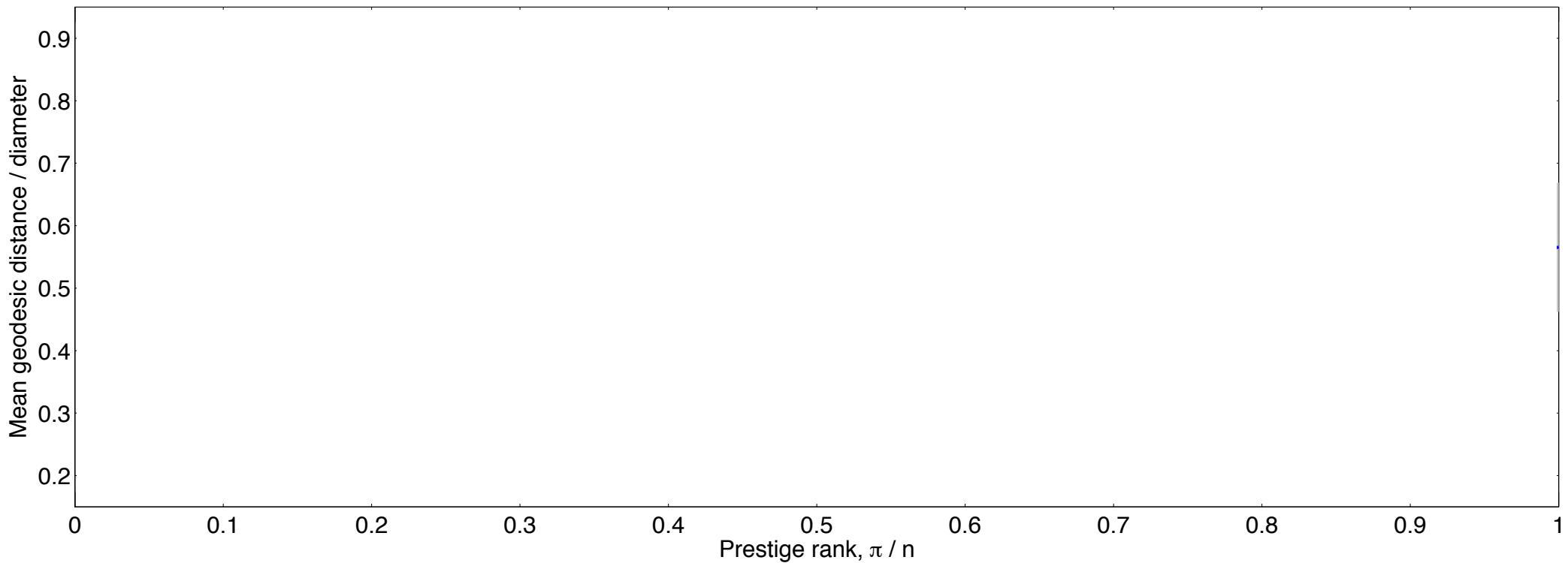
a rich club:

- top 15% of depts. produce 68% of their own faculty
- and hire only 7% from outside top 25% of depts.



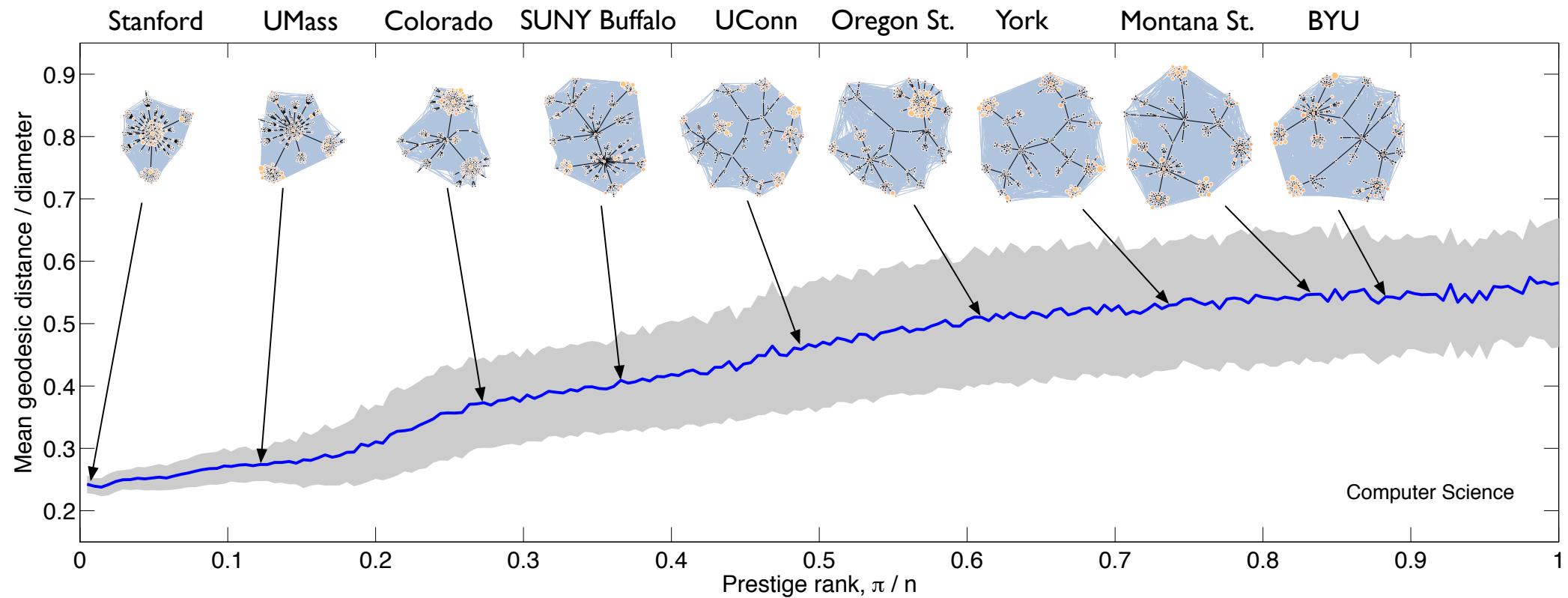
## architecture of faculty hiring networks

- how far is some  $v$  from a node  $u$ , on average?
- how does this distance vary with prestige?



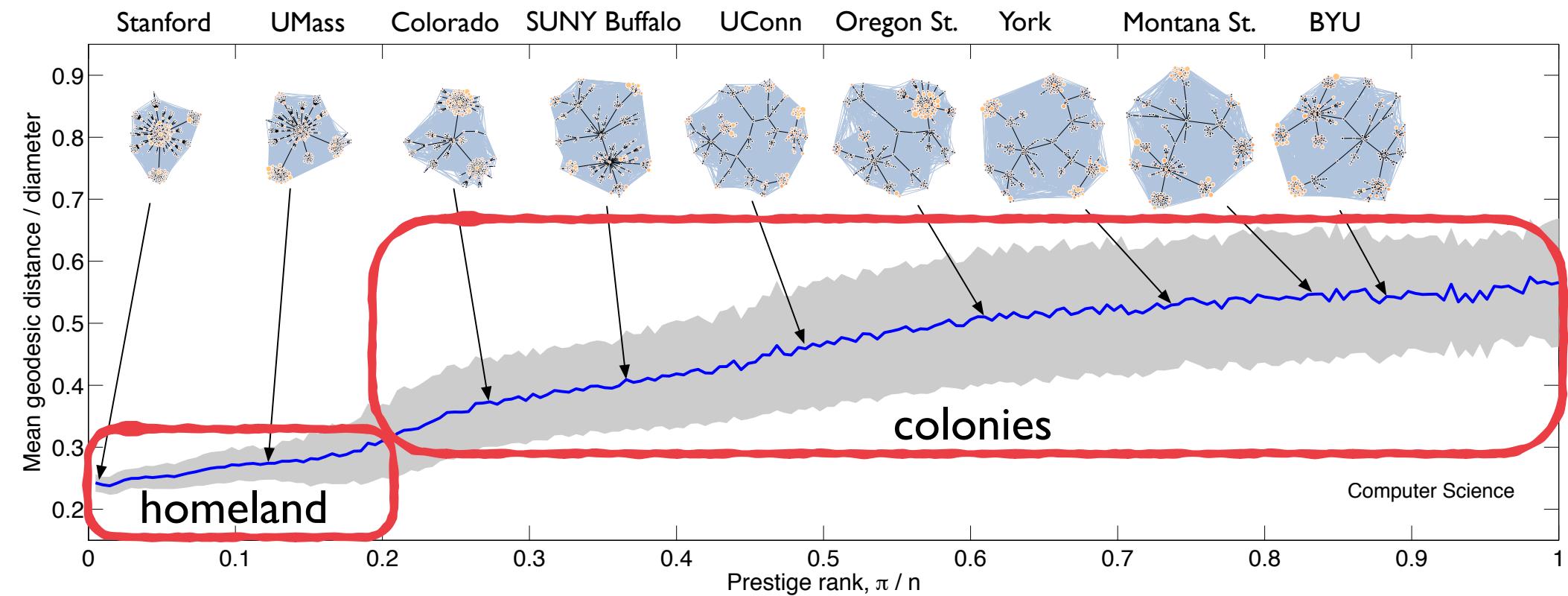
## architecture of faculty hiring networks

- core and periphery



## architecture of faculty hiring networks

- ~~core and periphery~~ *homeland and colonies*
- prestige is *influence*, via doctoral placement, over research agendas, research communities, and departmental norms across the discipline

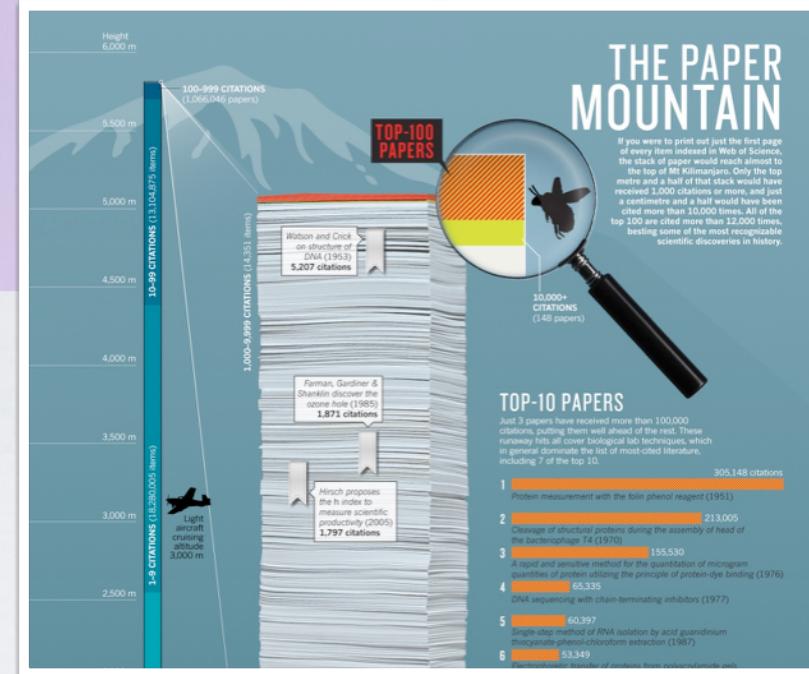


## prestige and faculty hiring networks

- prestige pervades and structures the academic ecosystem
  - placement power 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"
  - prestige is influence, via placement
- 🤔 how many discoveries are left unmade because hiring hierarchies so strongly determine who gets faculty positions?

# prestige drives epistemic inequality

some ideas spread further than others — why?



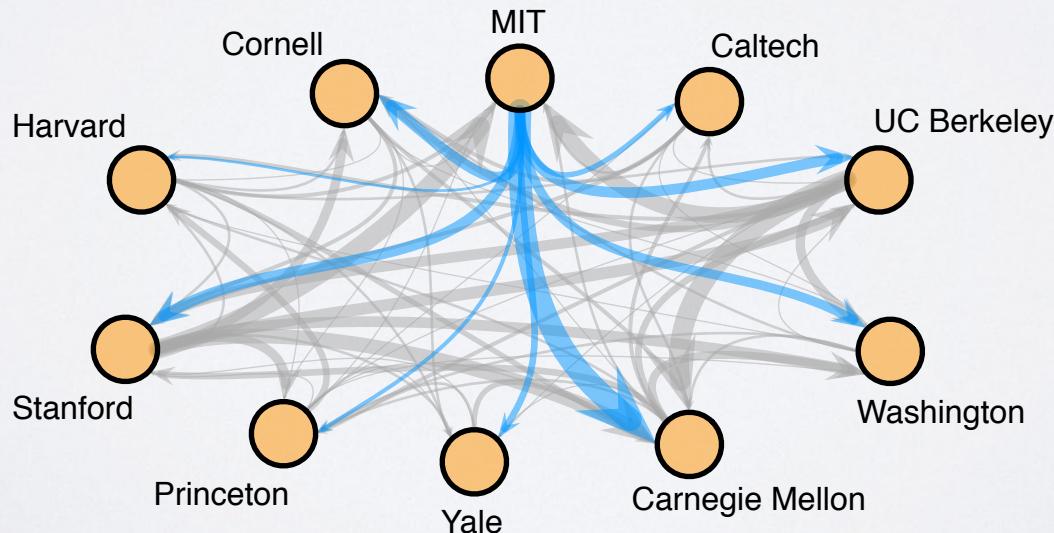
Nature "The top 100 papers." (2014)

# **prestige drives epistemic inequality**

1. genuine differences in merit } sorting is unbiased
2. non-meritocratic social processes } biases in who gets  
    ➤ fame, prestige, seniority, discrimination, history, etc.  
credit & opportunities
- 3.

# prestige drives epistemic inequality

1. genuine differences in merit } sorting is unbiased
2. non-meritocratic social processes } biases in who gets credit & opportunities
  - fame, prestige, seniority, discrimination, history, etc.
3. non-meritocratic structural factors } a mechanism:
  - scientists carry ideas from PhD to faculty institution
  - difference in placement power drives epistemic inequalitybiases in who does what work where  
agenda setting theory

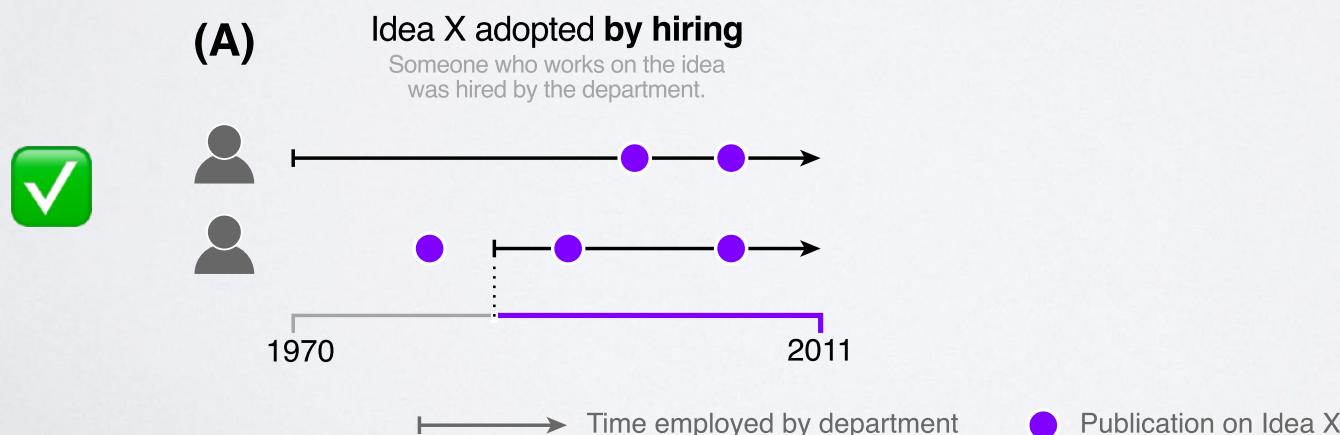


## test I: does faculty hiring drive spread of ideas?

### 3. non-meritocratic structural factors

- scientists carry ideas from PhD to faculty institution
- difference in placement power drives epistemic inequality

- using data on **hiring events & publication topics**, test whether **where an idea is being worked on** can be **explained by faculty placement** (and hence the prestige hierarchy)
- two possibilities:



# test I: does faculty hiring drive spread of ideas?

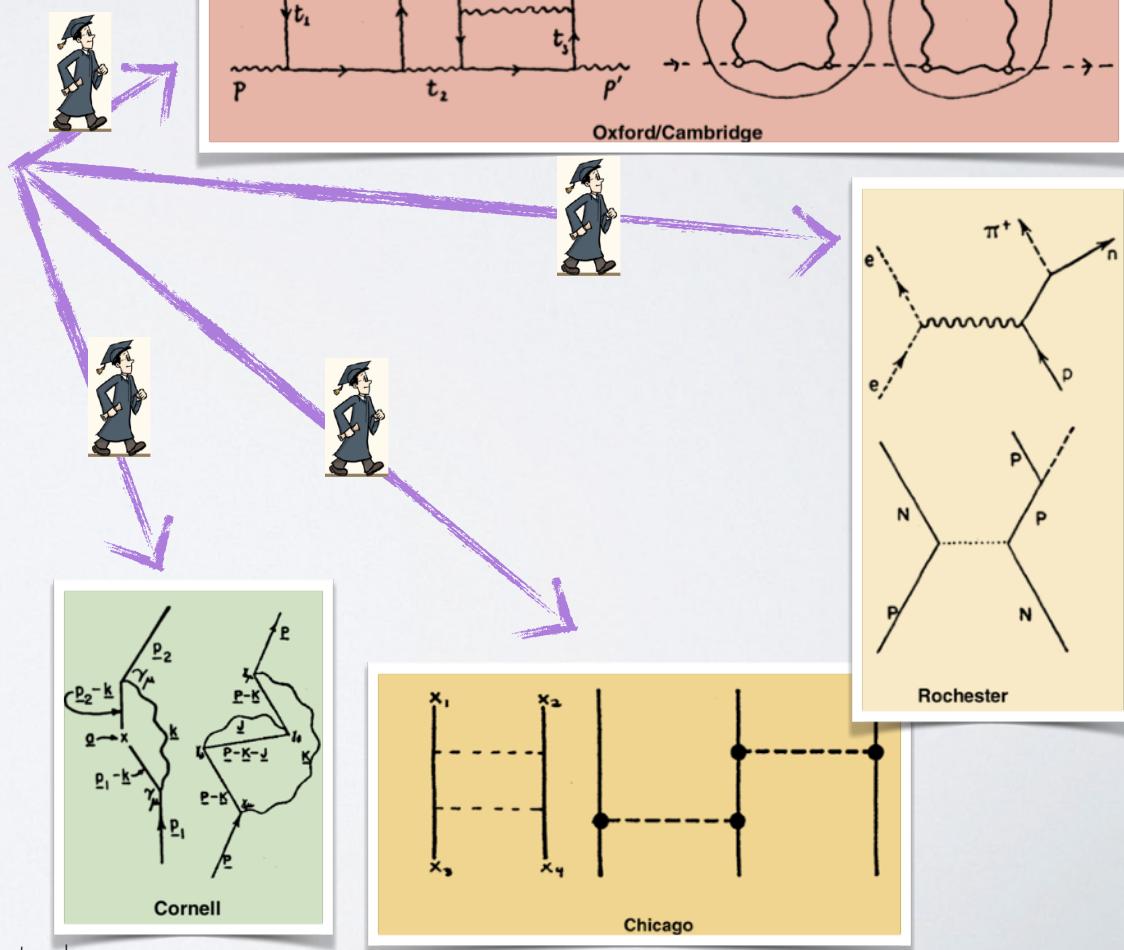
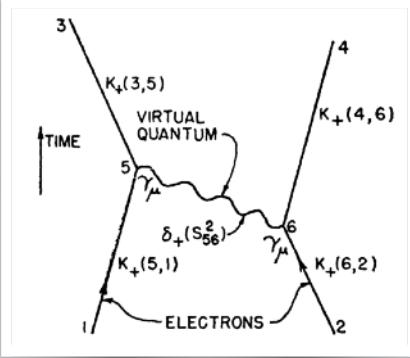
## 3. non-meritocratic structural factors

- scientists carry ideas
- for example

Feynman diagrams, born 1948



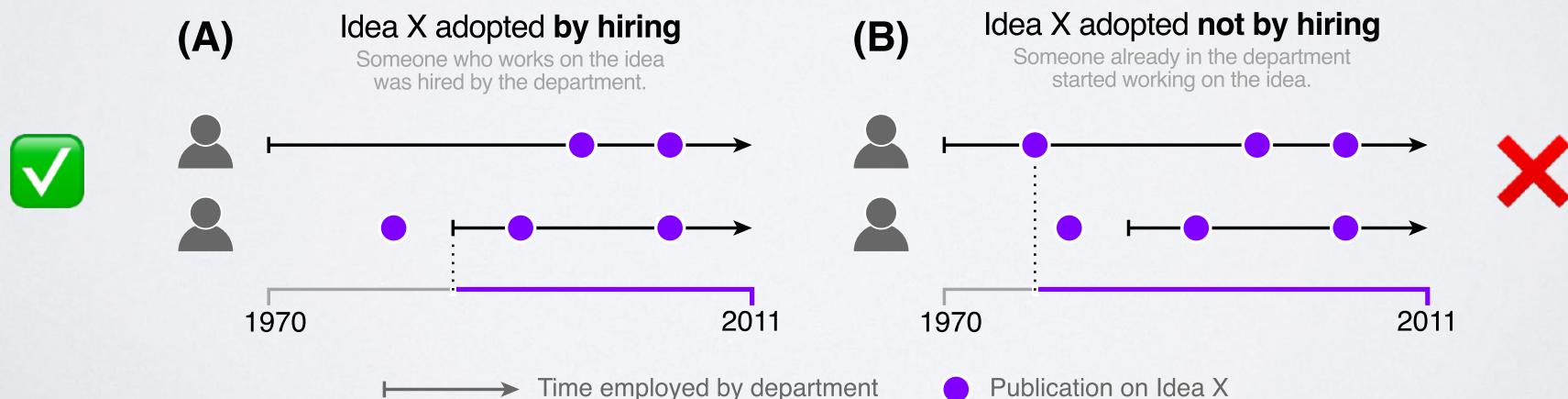
Lamb, Wheeler, Pais, Feynman,  
Feshbach & Schwinger



## test I: does faculty hiring drive spread of ideas?

### 3. non-meritocratic structural factors

- scientists carry ideas from PhD to faculty institution
- difference in placement power drives epistemic inequality
- using data on **hiring events & publication topics**, test whether **where an idea is being worked on** can be **explained by faculty placement** (and hence the prestige hierarchy)
- two possibilities:



test

Deep Learning

Incremental Computing

Topic Modeling

Up to 2000

Institutions arranged  
clockwise by  
prestige

Non-hiring  
adoption

+2

+3

(New adoption counts)

Adoption  
through hiring

test

Deep Learning

Incremental Computing

Topic Modeling

Up to 2000

Institutions arranged  
clockwise by  
prestige

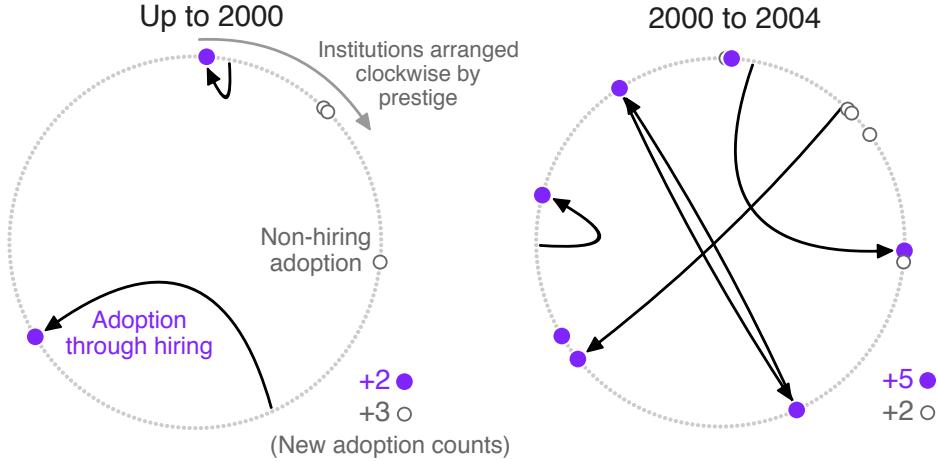
2000 to 2004

Non-hiring  
adoption

+2 ●  
+3 ○  
(New adoption counts)

Adoption  
through hiring

+5 ●  
+2 ○

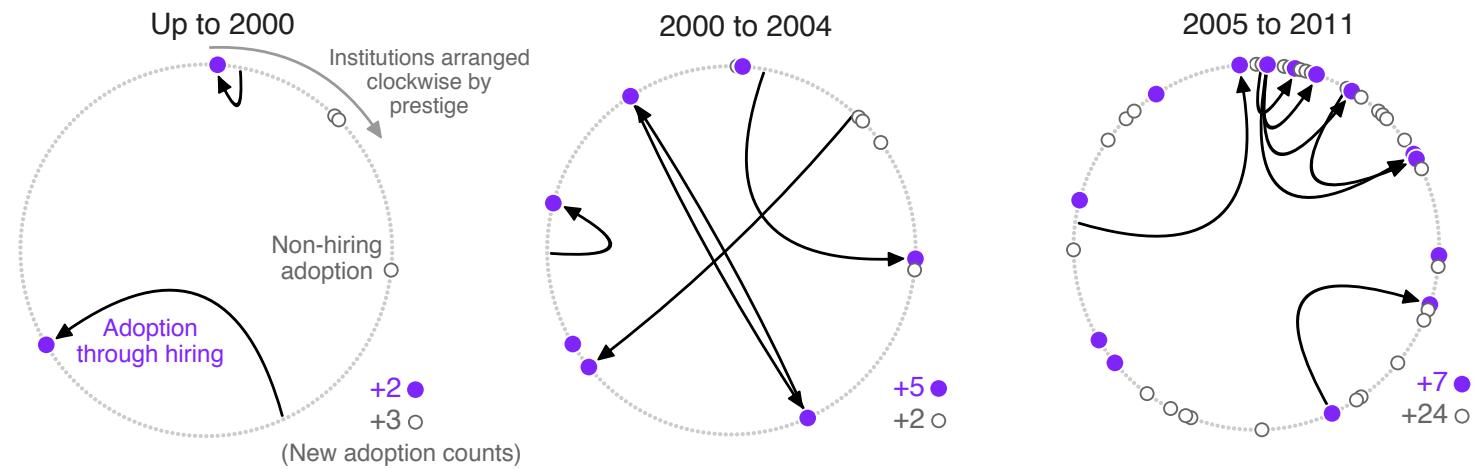


test

## Deep Learning

## Incremental Computing

## Topic Modeling



test

### Deep Learning

### Incremental Computing

### Topic Modeling

Up to 2000

Institutions arranged clockwise by prestige

2000 to 2004

Non-hiring adoption

Adoption through hiring

+2 ●  
+3 ○  
(New adoption counts)

2005 to 2011

+7 ●  
+24 ○

Up to 1990

+2 ●  
+1 ○

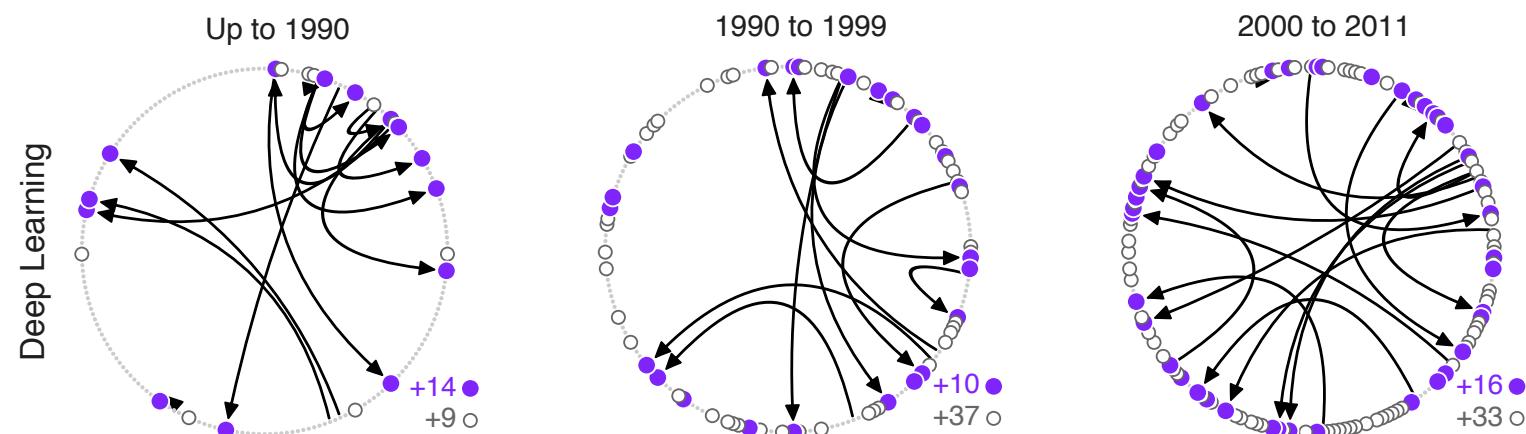
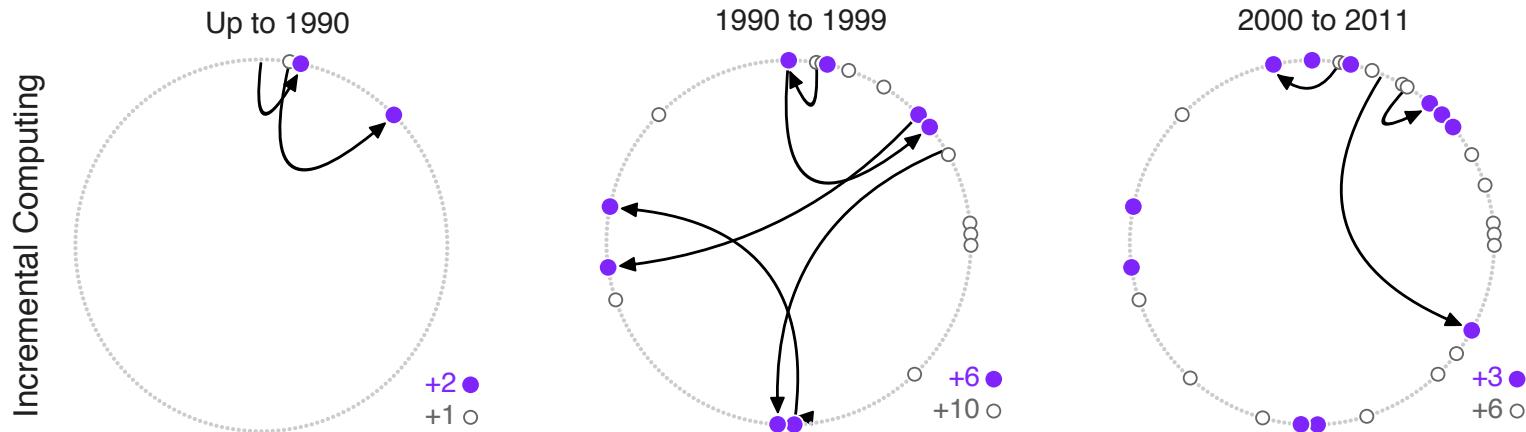
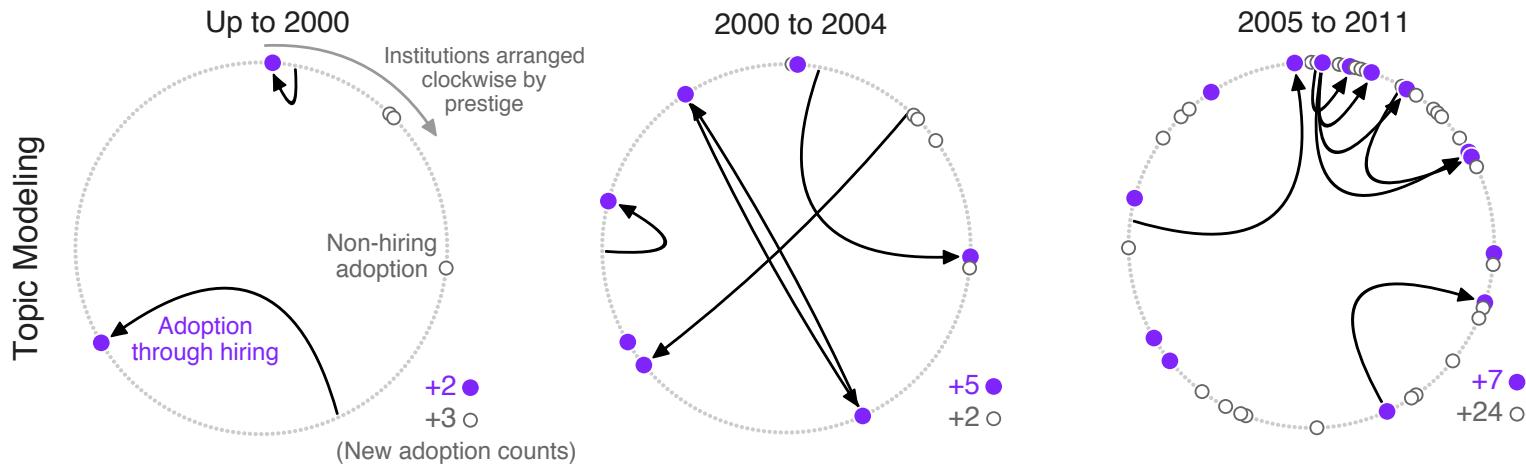
1990 to 1999

+6 ●  
+10 ○

2000 to 2011

+3 ●  
+6 ○

test



# test I: does faculty hiring drive spread of ideas?

topic X	$f_{\text{obs}}$	$f_{\text{exp}}$	$p$
topic modeling	0.35		
incremental computing	0.39		
deep learning	0.35		
quantum computing	0.32		
mechanism design	0.48		



fraction of real hires  
that spread topic X



fraction under  
permutation test

## test I: does faculty hiring drive spread of ideas?

topic X	$f_{\text{obs}}$	$f_{\text{exp}}$	$p$
topic modeling	0.35	0.23	$0.01 \pm 0.01$
incremental computing	0.39	0.20	$0.01 \pm 0.01$
deep learning	0.35	0.34	$0.34 \pm 0.01$
quantum computing	0.32	0.22	$0.01 \pm 0.01$
mechanism design	0.48	0.21	$0.01 \pm 0.01$



fraction of real hires  
that spread topic X



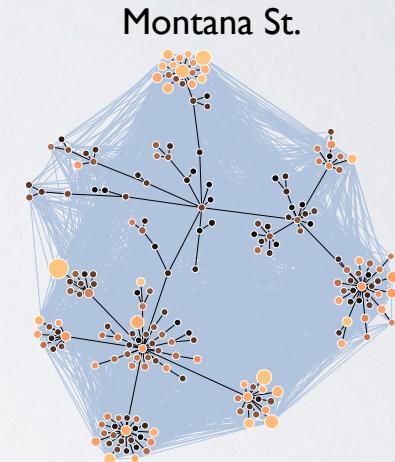
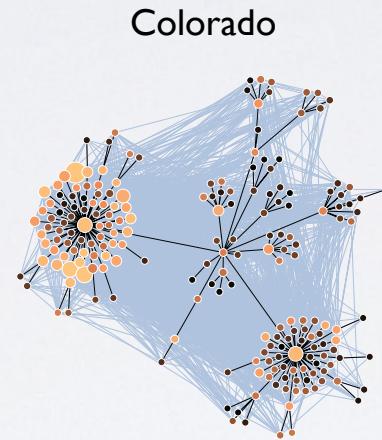
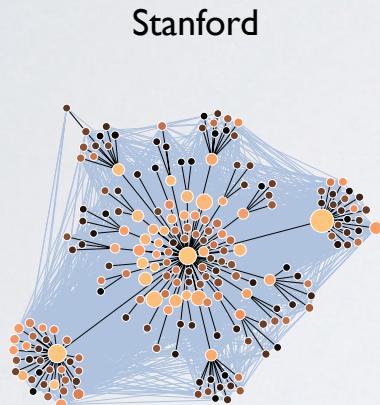
fraction under  
permutation test

→ **faculty hiring is a significant driver of the spread of ideas for some topics**

## test 2: spread ideas across faculty hiring network

how does point of origination shape spread of an idea?

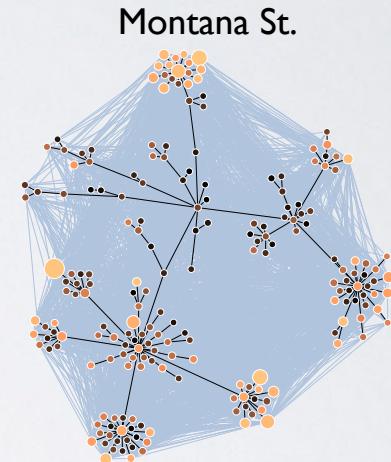
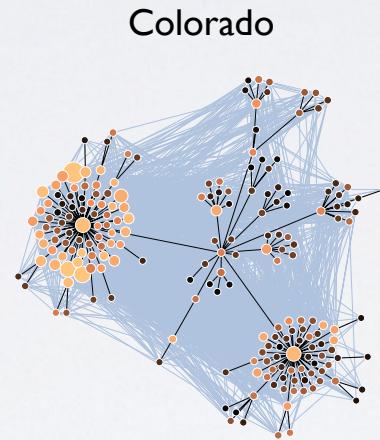
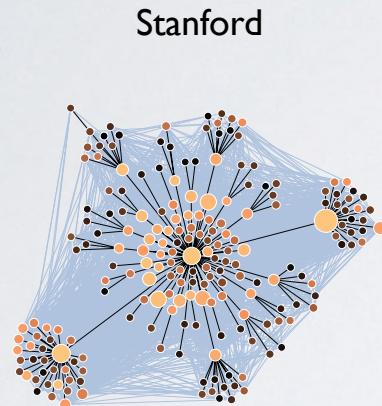
- prestige drives → core-periphery structure → drives spread



## test 2: spread ideas across faculty hiring network

how does point of origination shape spread of an idea?

- prestige drives → core-periphery structure → drives spread



model spread as Susceptible-Infected (SI) model on faculty hiring network

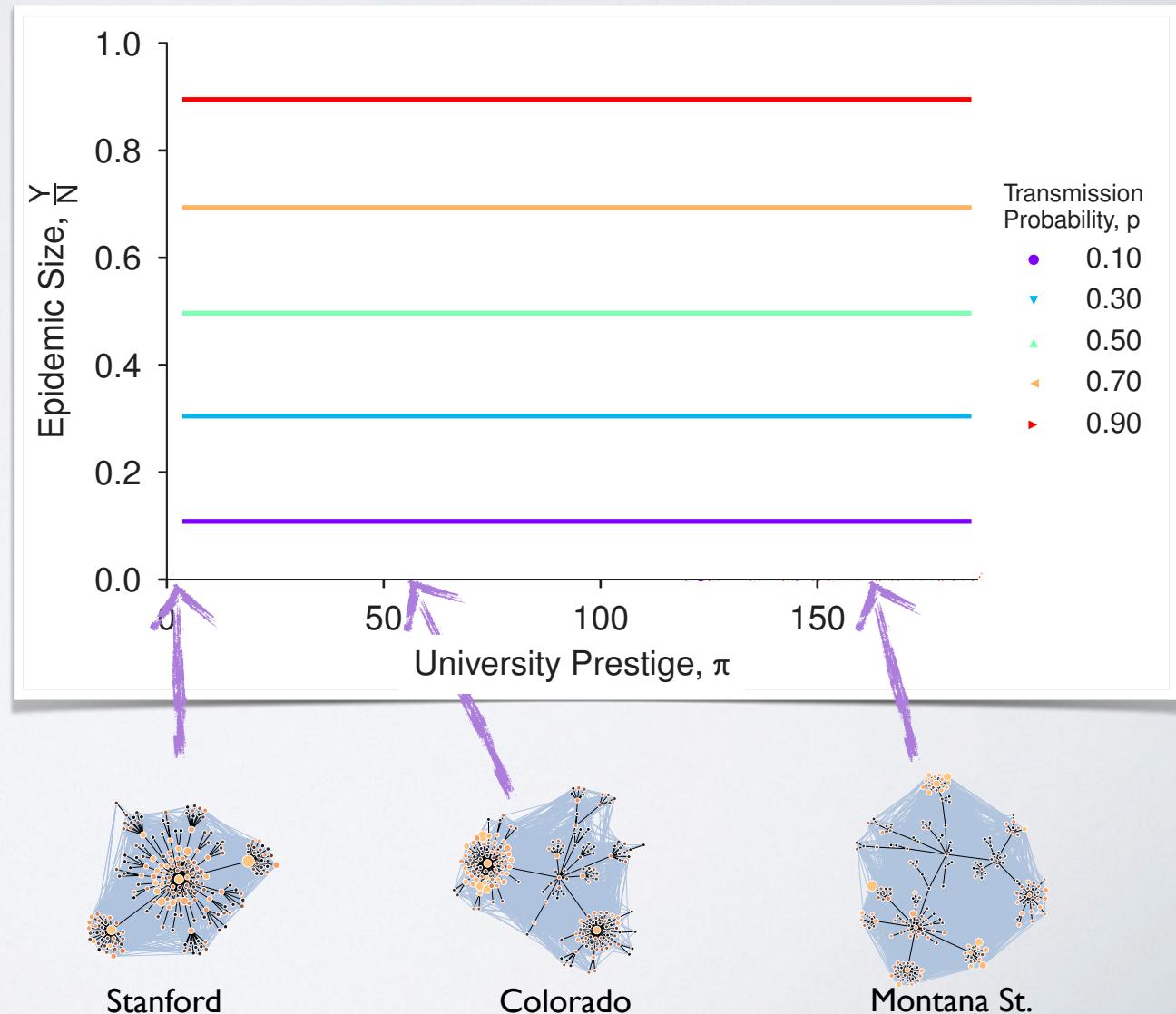
- seed simulation at each university, by prestige  $\pi$
- vary idea transmissibility  $p \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$
- measure normalized size  $Y/N$

## test 2: spread ideas across faculty hiring network

how does point of origination shape spread of an idea?

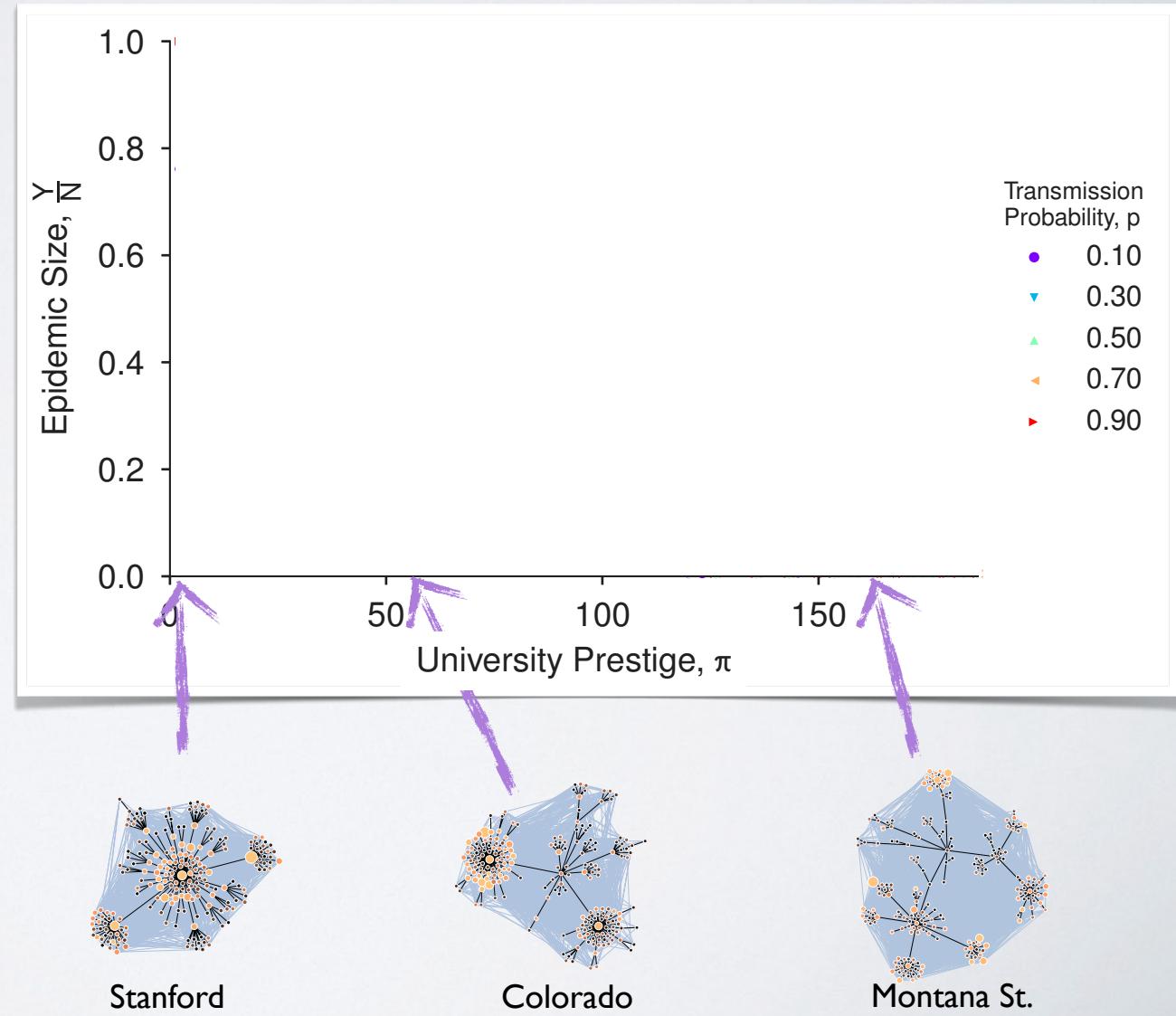
meritocracy?

- idea spread is proportional to idea "quality"
- spread is *independent* of birthplace



## test 2: spread ideas across faculty hiring network

how does point of origination shape spread of an idea?

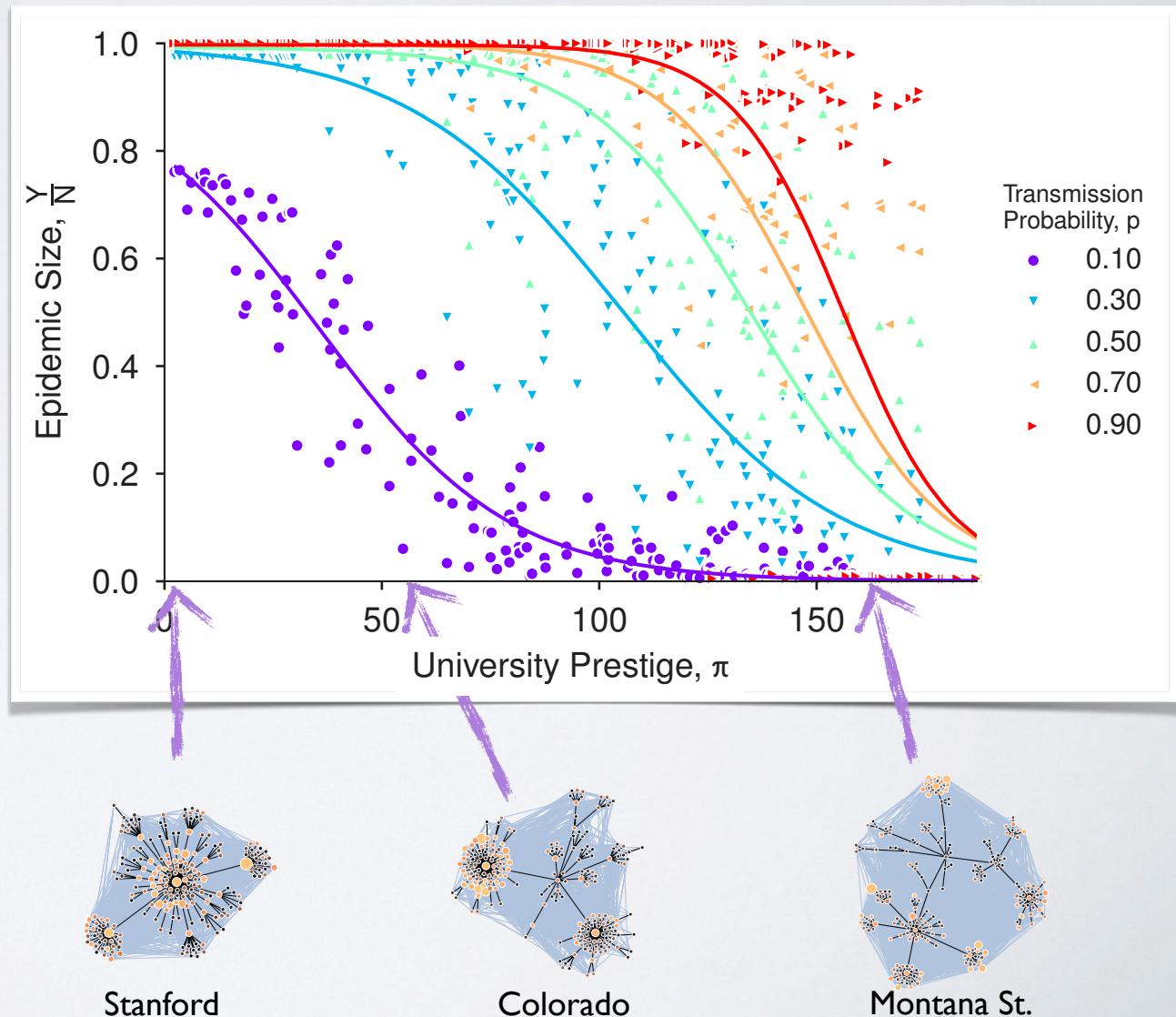


## test 2: spread ideas across faculty hiring network

how does point of origination shape spread of an idea?

not-a-meritocracy.

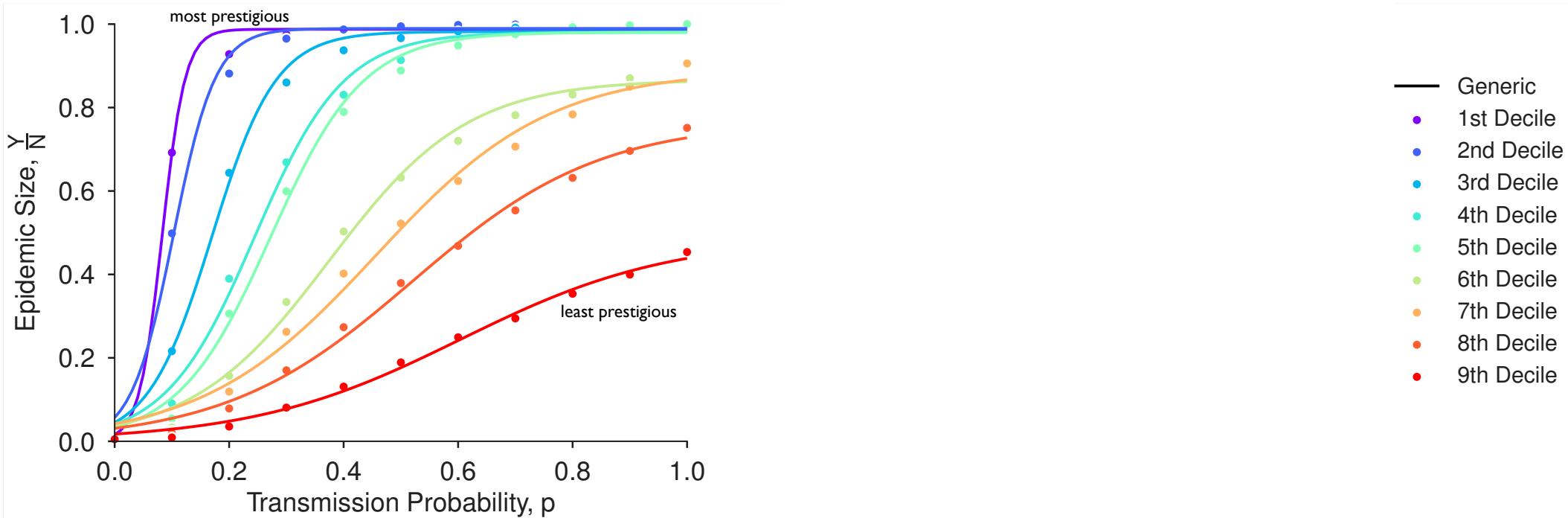
- 👍 • best ideas spread independent of origin
- 👎 • bad ideas spread easily from elite departments
- 👎 • good ideas from mid-prestige spread less well than bad ideas from high-prestige



## **test 2: spread ideas across faculty hiring network**

how does point of origination shape spread of an idea?

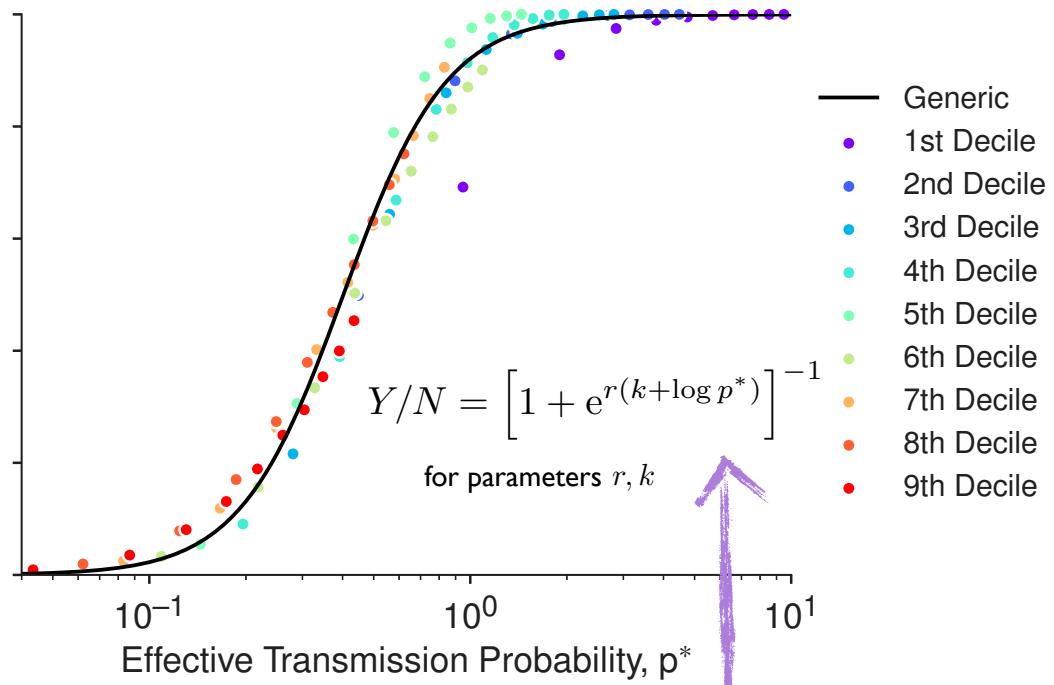
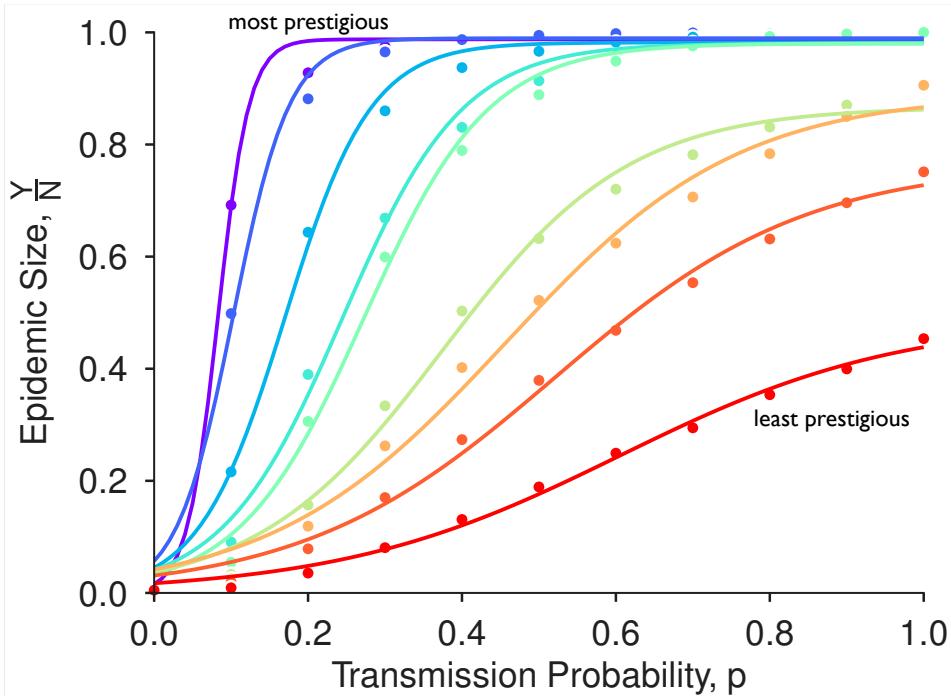
- classic logistic growth  $\rightarrow$  higher  $p$ , bigger spread  $Y/N$
  - rescale  $p^* = -p / \log(1 - d)$  to collapse curves for prestige deciles  $d$



## test 2: spread ideas across faculty hiring network

how does point of origination shape spread of an idea?

- classic logistic growth  $\rightarrow$  higher  $p$ , bigger spread  $Y/N$
- rescale  $p^* = -p / \log(1 - d)$  to collapse curves for prestige deciles  $d$



prestige is an attention amplifier

## epistemic inequalities

3. inequalities reflect non-meritocratic structural factors

- ✓ scientists carry ideas from PhD to faculty institution
  - ✓ difference in placement power drives epistemic inequality
- 

prestige → placement → spread of ideas

- simulations suggests an exponential dependence of "impact" on increased prestige
- faculty hiring network is a *structural mechanism*: it creates & maintains epistemic inequality
- your location in network shapes how far your ideas spread

## **networks, prestige, and the spread of scientific ideas**

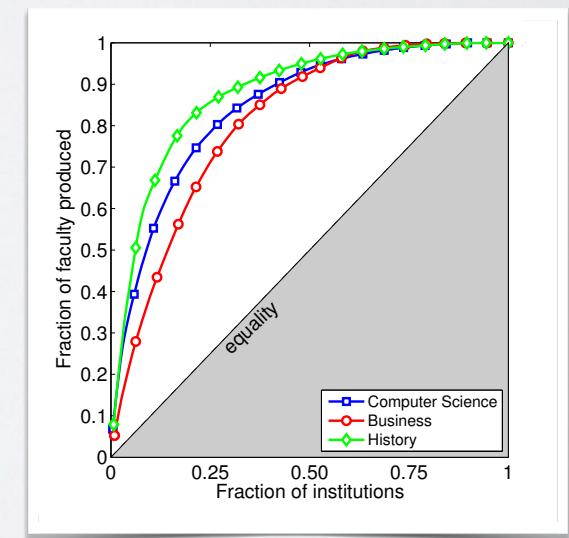
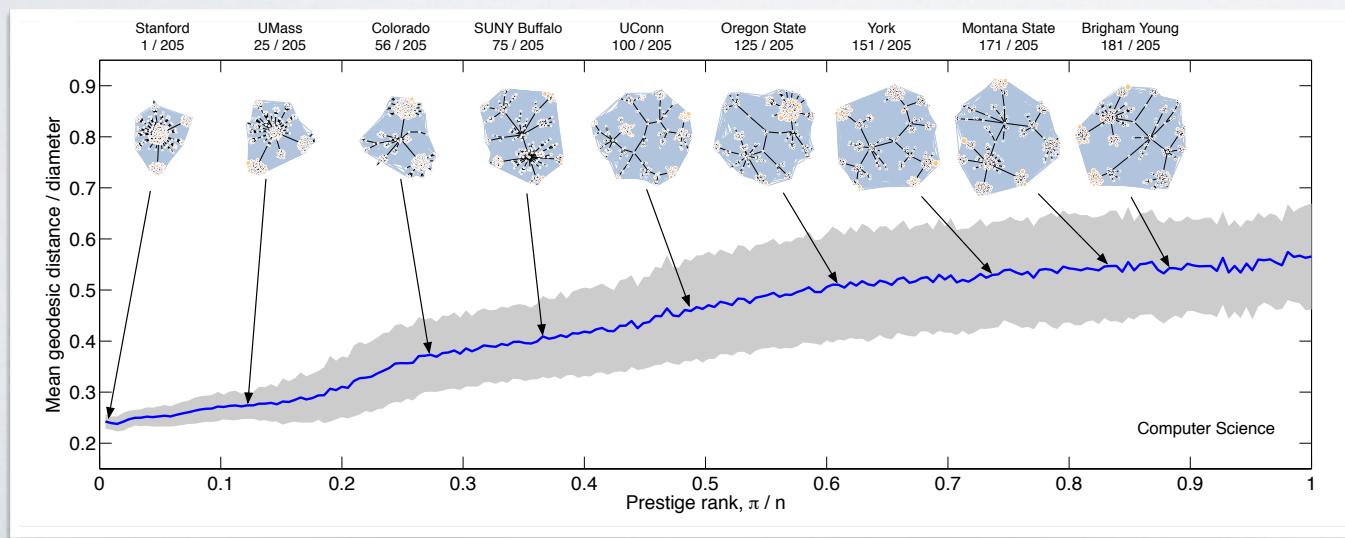
*prestige pervades and structures the scientific ecosystem*

- jobs, publications, spread of ideas, resources...

# networks, prestige, and the spread of scientific ideas

prestige pervades and structures the scientific ecosystem

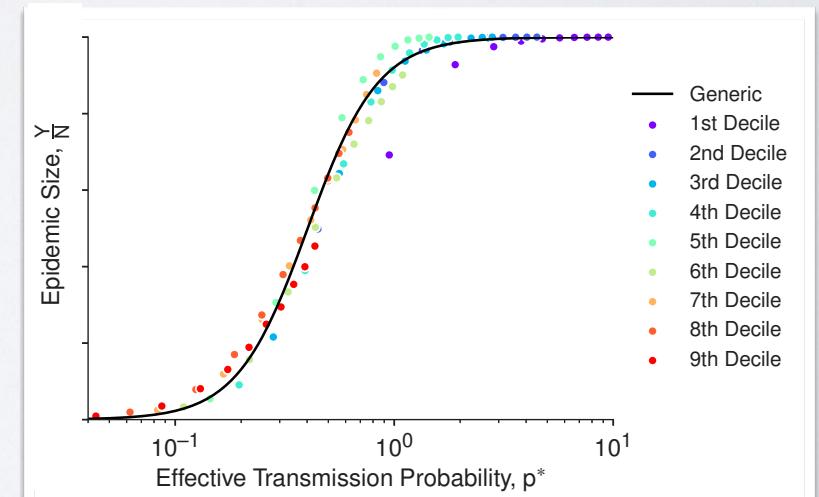
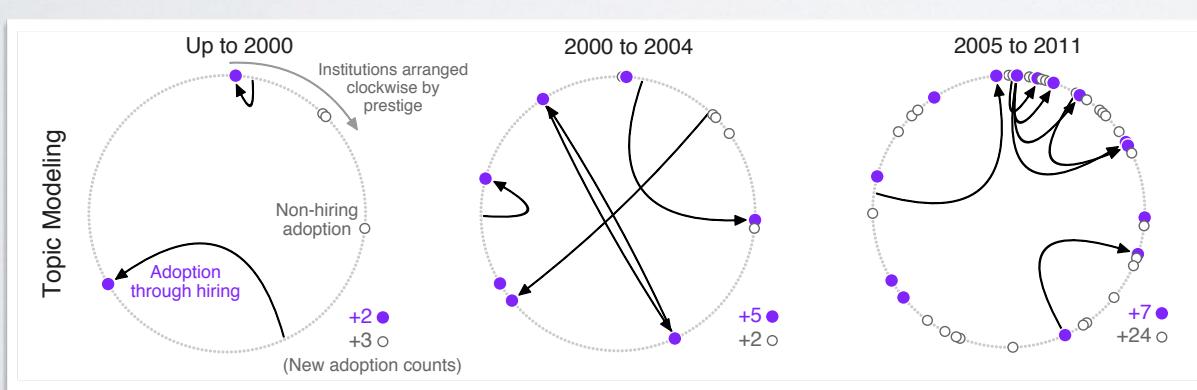
- jobs, publications, spread of ideas, resources...
- hiring networks → core-periphery structure (homeland/colonies)
  - highly-skewed faculty production
  - latent one-D embedding (a hierarchy)



# networks, prestige, and the spread of scientific ideas

prestige pervades and structures the scientific ecosystem

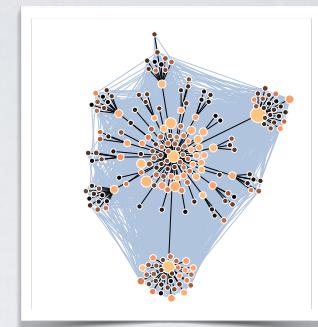
- jobs, publications, spread of ideas, resources...
- hiring networks → drives spread of ideas
  - prestige is an idea amplifier
  - non-meritocratic bias



# **networks, prestige, and the spread of scientific ideas**

*prestige pervades and structures the scientific ecosystem*

- jobs, publications, spread of ideas, resources...
- hiring networks → prestige.



*science is a model complex system*

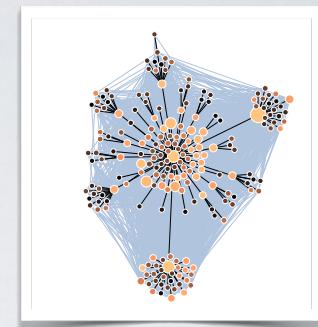
- competition, cooperation, feedback, incentives, networks, multi-scales
- hypothesis: scientific discovery is an emergent property of the system

big questions:

# **networks, prestige, and the spread of scientific ideas**

*prestige pervades and structures the scientific ecosystem*

- jobs, publications, spread of ideas, resources...
- hiring networks → prestige



*science is a model complex system*

- competition, cooperation, feedback, incentives, networks, multi-scales
- hypothesis: scientific discovery is an emergent property of the system

(big questions:

- does prestige help or hinder scientific progress?
- how much competition is too much or too little?
- what changes would accelerate discoveries?
- what discoveries are not being made because of system bias?

## NETWORK SCIENCES

# Systematic inequality and hierarchy in faculty hiring networks

Aaron Clauset,<sup>1,2,3\*</sup> Samuel Arbesman,<sup>4</sup> Daniel B. Larremore<sup>5,6</sup>

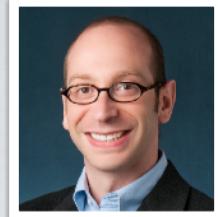
Science Advances 1, e1400005 (2015)



Prof. Aaron Clauset  
(Colorado)



Prof. Daniel Larremore  
(Colorado)



Dr. Sam Arbesman  
(now Lux Capital)

# Prestige drives epistemic inequality in the diffusion of scientific ideas

Allison C. Morgan<sup>1\*</sup>, Dimitrios J. Economou<sup>1</sup>, Samuel F. Way<sup>1</sup> and Aaron Clauset<sup>1,2,3</sup>

EPJ Data Science 7, 40 (2018)



Dr. Allison Morgan  
(now Twitter)



Dimitrios Economou  
(Queen's U.)



Dr. Sam Way  
(now Spotify)

# Productivity, prominence, and the effects of academic environment

Samuel F. Way<sup>a,1</sup>, Allison C. Morgan<sup>a</sup>, Daniel B. Larremore<sup>a,b,2</sup>, and Aaron Clauset<sup>a,b,c,1,2</sup>

<sup>a</sup>Department of Computer Science, University of Colorado, Boulder, CO, USA; <sup>b</sup>BioFrontiers Institute, University of Colorado, Boulder, CO, USA; <sup>c</sup>Santa Fe Institute, Santa Fe, NM, USA

Proc. Natl. Acad. Sci. USA 116, 10729 (2019)

## ESSAY

# Data-driven predictions in the science of science

Aaron Clauset,<sup>1,2,\*</sup> Daniel B. Larremore,<sup>2</sup> Roberta Sinatra<sup>3,4</sup>

Science 355, 477 (2017)



Prof. Roberta Sinatra  
(ITU Copenhagen)



# fin

see also:

Morgan et al., "Socioeconomic roots of academic faculty." Preprint (2021)

Morgan et al. "The unequal impact of parenthood in academia." *Science Advances* 7 (2021)

Way et al., "Productivity, prominence, and the effects of academic environment." *PNAS* 116 (2019)

Morgan et al., "Prestige drives epistemic inequality in the diffusion of scientific ideas." *EPJ Data Science* 7 (2018)

Way et al., "The misleading narrative of the canonical faculty productivity trajectory." *PNAS* 114 (2017)

Clauset et al., "Data-driven predictions in the science of science." *Science* 355 (2017)

Way et al., "Gender, productivity, and prestige in computer science faculty hiring networks." *Proc. WWW* (2016)

Clauset et al., "Systematic inequality and hierarchy in faculty hiring networks." *Science Advances* 1 (2015)



## prestige shapes productivity — why?

- ▶ scientific discourse is dominated by elite scientists and elite institutions

but:





## prestige shapes productivity — why?

- ▶ scientific discourse is dominated by **elite scientists** and **elite institutions**

but:



- ▶ this dynamic reflect **endogenous cumulative advantage** : past achievements correlate with future achievements



what causes differences in individual or institutional scholarly achievement?



## prestige shapes productivity — why?

▶ what drives future productivity and prominence?

### idea I: where a scientist trained.

- skill, talent, training, temperament, etc.
- faculty hiring market sorts individuals by their natural or previously acquired characteristics that correlate with outcomes
- e.g., most productive scientists have elite pedigree, Harvard, Yale, Penn, Stanford, etc.

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



## prestige shapes productivity — why?

▶ what drives future productivity and prominence?

**idea 1:** where a scientist trained.

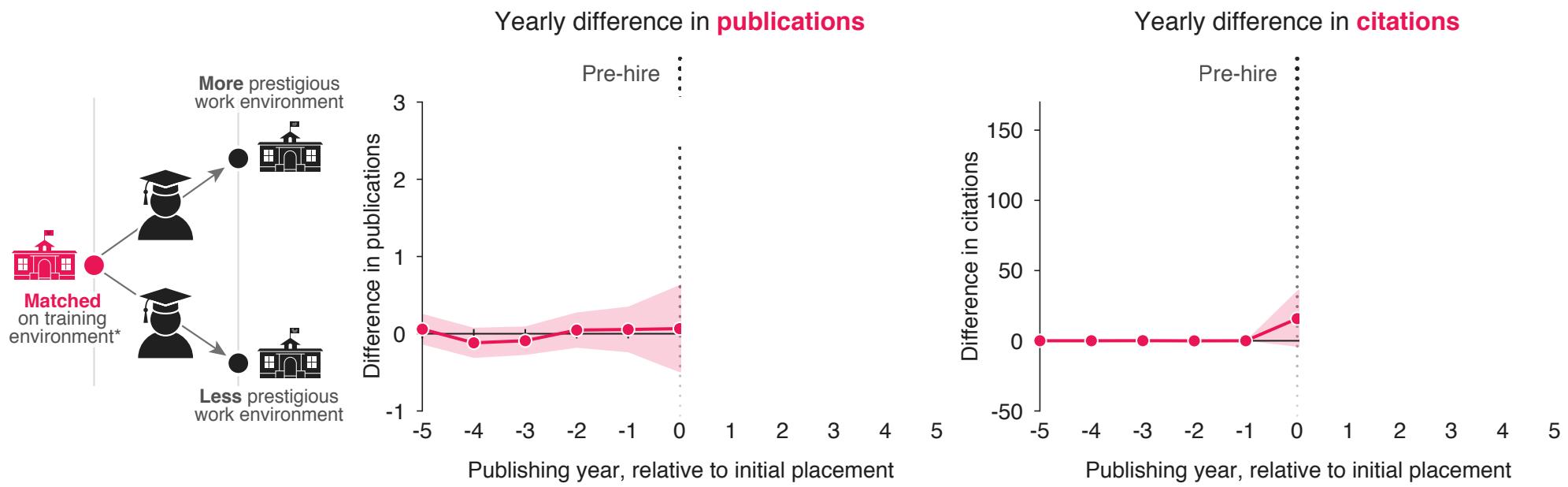
**idea 2:** where a scientist works.

- 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 (Allison & Long 1990)

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

# quasi-natural experiments : faculty hiring

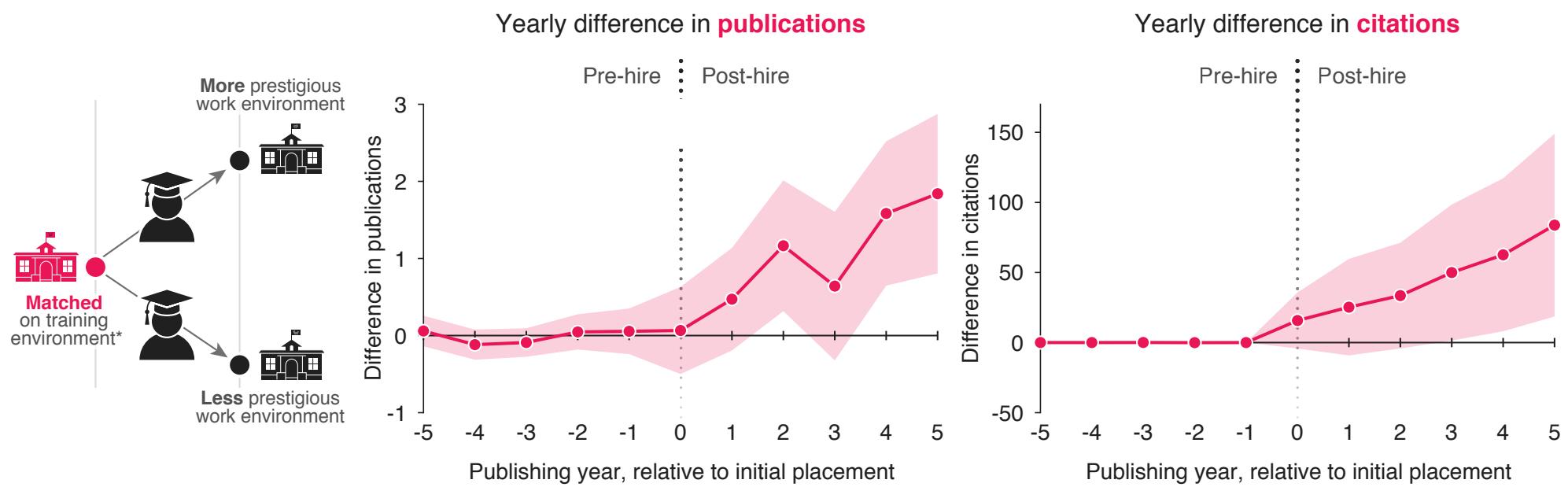
similar training, different placement



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  
citations:  $N = 96$  pairs

# quasi-natural experiments : faculty hiring

similar training, different placement → different productivity & prominence



publications:  $N = 196$  pairs  
citations:  $N = 96$  pairs

## **prestige shapes productivity — why?**

*conditioned on holding a faculty position:*

**more elite training does not drive greater scholarly impact  
relative to peers in similar faculty positions**

work environment, not training appears to drive impact

## **prestige shapes productivity — why?**

*conditioned on holding a faculty position:*

**more elite training does not drive greater scholarly impact  
relative to peers in similar faculty positions**

work environment, not training appears to drive impact

⭐ why do elite institutions dominate science?

doctoral prestige → faculty location

Clauset et al. (2015), Way et al. (2016)

## prestige shapes productivity — why?

*conditioned on holding a faculty position:*

**more elite training does not drive greater scholarly impact  
relative to peers in similar faculty positions**

work environment, not training appears to drive impact

⭐ why do elite institutions dominate science?

doctoral prestige → faculty location → scholarly impact

these results, Way et al. (2019)

why? because of **working environment**

[not selection, not retention, not expectations]

## prestige shapes productivity

- ★ where you work is an *environmental mechanism* to explain cumulative advantage
- ★ individual productivity and prominence *cannot be separated* from their place in the academic system

### future work:

how does this effect vary across different fields? why?

what department-level *facilitation* mechanisms drive productivity?

what role for differences in students?