

# Mentorship networks and the science of science

Stephen David

*davids@ohsu.edu*

28 July 2021

# Why study academic mentorship?

- Most researchers spend several years training with just one or two mentors.
- Trainees learn numerous theoretical and practical skills from their mentor.
- Science, like any other field, is a social network. A recommendation is worth a lot.
- It's also a very practical aspect of your career!

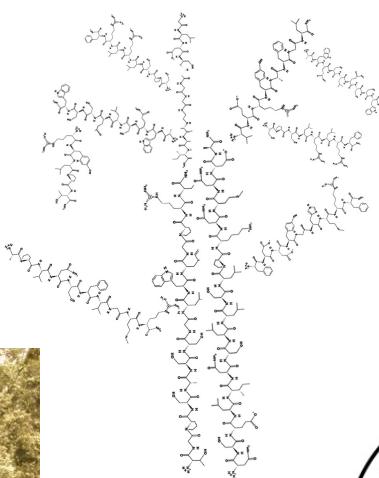
# Data available on scientists

- Mentor “genealogy”
  - Who trains with who (PhD student, Postdoc)
  - When?
- Publication
  - Number of publications
  - Publication content (topics)
  - Co-authorship
  - Number of citations & h-index
- Funding levels
- Awards (example: Nobel prize)
- Geography
  - Institute size & renown
  - Country
- Demographics
  - Gender
  - Age (more precisely: “academic age”)

# Data: academic genealogy



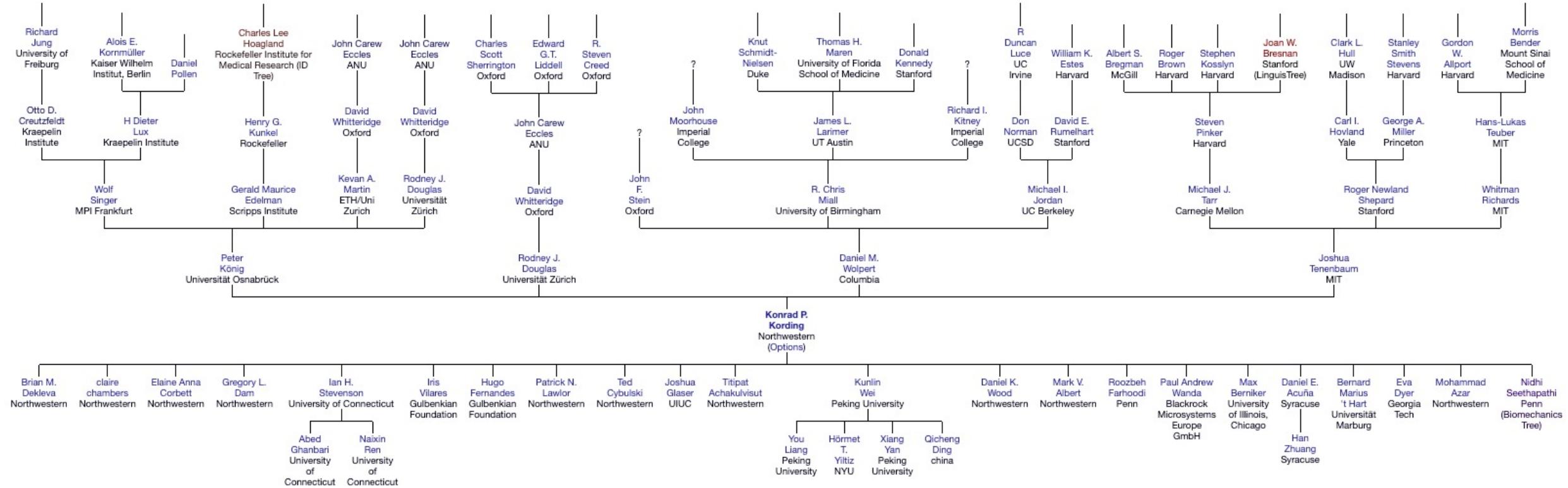
Neuro Tree



## The Academic Family Tree

- Crowdsourced database of mentoring relationships (<https://academictree.org>)
- Presented in a “family tree” format.
- ~100 fields, ~800K researchers
- Data export available on soshub (~/datasets/s4/MENTORSHIP/)
- or at  
<https://zenodo.org/record/4441298#.YPc8bxNKj0o>

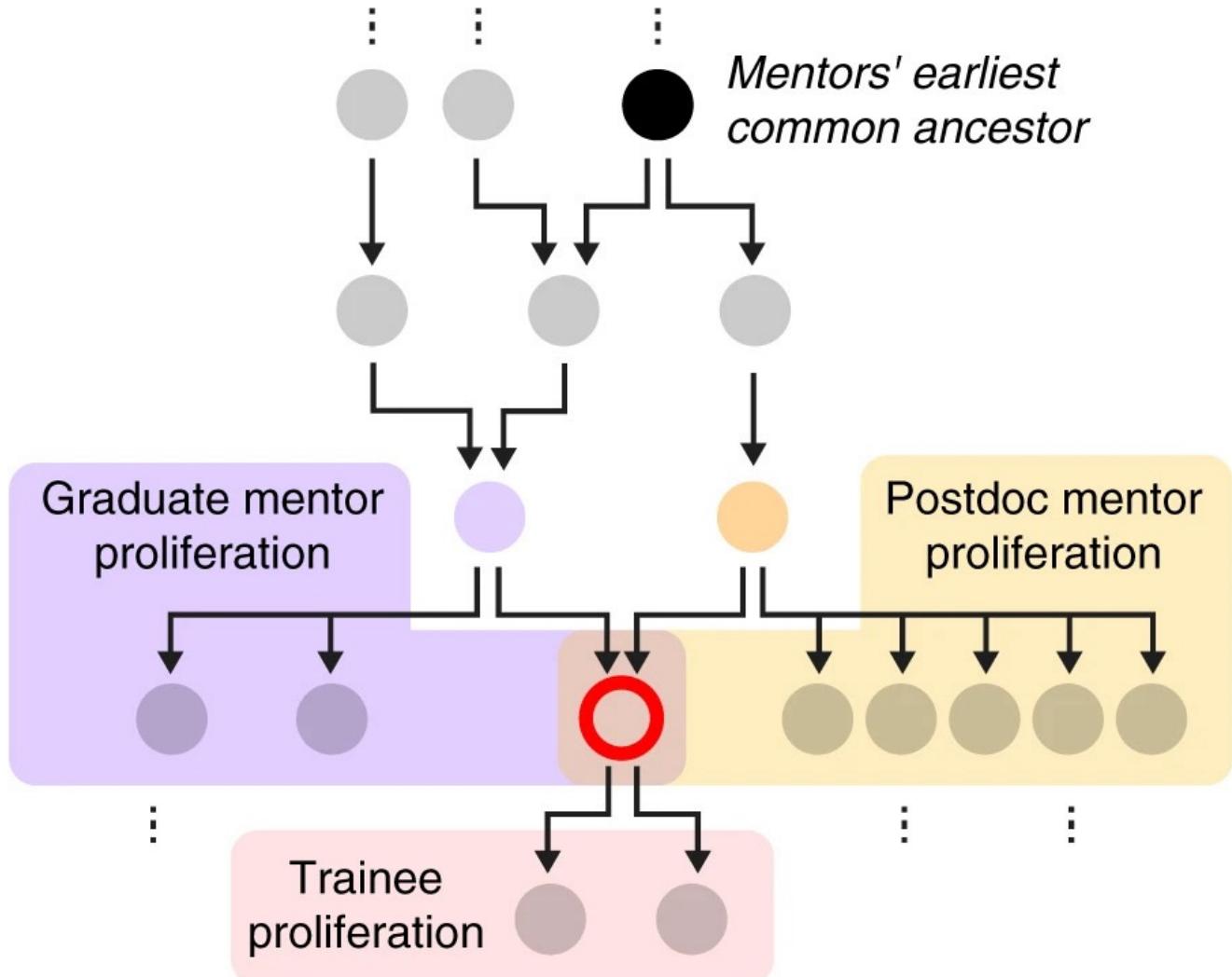
# Data: academic genealogy



# Data: academic genealogy

## Graph-based metrics

- *Proliferation rate* measures number of trainees as a proxy for professional success
- *Common ancestor distance* measures steps between two nodes through a shared ancestor



# Data: publications

Focus on traditional publications

- Data from Pubmed, Semantic Scholar, and Microsoft Academic Graph
- The challenge is matching author names to network nodes unambiguously (thanks Daniel!)



## Konrad P. Kording

Affiliations: Northwestern University, Evanston, IL (Edit/merge)  
University of Pennsylvania, Philadelphia, PA, United States (Edit/merge)  
Area: Motor Control, Bayesian statistics, Normative models  
Website: <http://www.koerding.com/>  
Google: "Konrad Kording"  
Mean distance: 12.84 (update) (cluster 29) SNECH  
Cross-listing: MathTree (X) - Computer Science Tree (X) - Physics Tree (X) - Add to: --  
Edit user koerding

Tree Publications Grants Sim nodes Sim pubs Dist Nobel Influence Reading PubMed Edit Del Rollback Error

Parents					New mentor
Peter König	grad student	Universität Zürich	(hanks)	Delete Edit	
Rodney J. Douglas	grad student	1999-2003	Universität Zürich	(koerding)	Delete Edit
Daniel M. Wolpert	post-doc	2002-2004	UCL	(david)	Delete Edit
Joshua Tenenbaum	post-doc	2004-2006	MIT	(david)	Delete Edit

Children					New trainee
claire chambers	grad student	Northwestern	(koerding)	Delete Edit	
Brian M. Dekleva	grad student		(bdekleva)	Delete Edit	
Patrick N. Lawlor	grad student	2012-	Northwestern	(pnlawlor)	Delete Edit
Gregory L. Dam	grad student	2006-2009	Northwestern	(pq)	Delete Edit
Ian H. Stevenson	grad student	2006-2011	Northwestern	(pq)	Delete Edit
Elaine Anna Corbett	grad student	2012	Northwestern	(pq)	Delete Edit

Topics

physiology (65) Humans (58) Models, Neurological (28) Male (28) Animals (27) Female (27) Movement (27) Psychomotor Performance (23) Bayes Theorem (21) Adult (21) Learning (17) Neurons (17) Algorithms (16) Photic Stimulation (15) methods (15)

Publications

You can help our author matching system! Click on the check if a publication is attributed accurately or the "x" if it is not.

Achakulvisut T, Ruangrong T, Mineault P, et al. (2021) Towards Democratizing and Automating Online Conferences: Lessons from the Neuromatch Conferences. *Trends in Cognitive Sciences* ✓ ? ✗

Chambers C, Seethapathi N, Saluja R, et al. (2020) Computer vision to automatically assess infant neuromotor risk. *Ieee Transactions On Neural Systems and Rehabilitation Engineering : a Publication of the Ieee Engineering in Medicine and Biology Society* ✓ ? ✗

Naidech AM, Lawlor PN, Xu H, et al. (2020) Probing the Effective Treatment Thresholds for Alteplase in Acute Ischemic Stroke With Regression Discontinuity Designs. *Frontiers in Neurology*. 11: 961 ✓ ? ✗

Glaser JI, Benjamin AS, Chowdhury RH, et al. ✓

# Data: publications

## Identifying/comparing semantic content

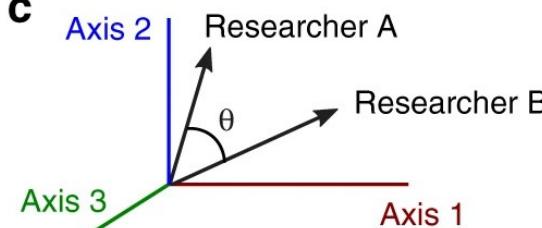
- Abstracts are easily and freely accessible
- Latent Semantic Analysis (TF-IDF + PCA) to generate vector space.  
(\*\* Note that SPECTER embeddings now available. \*\*)
- Cosine similarity to compare publications (scaled to range 0-10)
- More recently measuring SPECTER embedding and identifying MAG keywords

**b**

Synap -se/ses/tic  
Neurotransmitter -s  
Neuron -s/al  
Parkinson -ian  
Dopamine -rgic  
Striatum  
Rat -s  
Rodent -s  
Murine  
⋮

	Trainee		Graduate mentor		Postdoc mentor	
Synap -se/ses/tic	2	0	0	1	0	0 0 1
Neurotransmitter -s	0	1	0	0	0	1 1 1
Neuron -s/al	0	1	0	0	0	0 1 0
Parkinson -ian	0	2	0	1	2	1 1 1
Dopamine -rgic	1	3	0	0	0	0 0 1
Striatum	0	0	0	1	1	0 0 1
Rat -s	1	1	2	1	0	0 1 0
Rodent -s	1	0	0	0	0	0 0 0
Murine	0	0	1	0	0	0 0 1
⋮	⋮	⋮	⋮	⋮	⋮	⋮

**c**



Document vector  
Term vector

# Data: funding

- Open data, currently from US federal agencies only (Federal Reporter)
  - NIH
  - NSF
  - NASA
- Linked using similar disambiguation procedure as for publications.
- Notably missing data from Department of Defense, private foundations, and other countries.

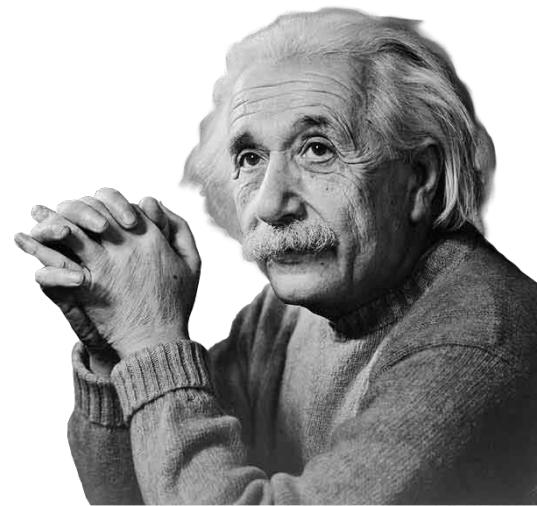
High-probability grants					
According to our matching algorithm, Konrad P. Kording is the likely recipient of the following grants.					
Filter high-probability grants:					
Years	Recipients	Code	Title / Keywords	Matching score	
2009 – 2013	Kording, Konrad P.	R01	The Role of Uncertainty in Human Motor Learning and Adaptation @ Rehabilitation Institute of Chicago <a href="#">Show summary</a>	0.892	
2009 – 2013	Kording, Konrad P.	P01	Bayesian and Decision Theoretic Tools @ University of California Santa Barbara <a href="#">Show summary</a>	0.942	
2009 – 2015	Miller, Lee Tresch, Matthew Lynch, Kevin Perreault, Eric <b>Kording, Konrad (co-PI)</b>	N/A	Cps: Large: Cybernetic Interfaces For the Restoration of Human Movement Through Functional Electrical Stimulation @ Rehabilitation Institute of Chicago <a href="#">Show summary</a>	0.807	
2010 – 2014	David, Ostry Miller, Lee <b>Kording, Konrad</b> Thoroughman, Kurt	N/A	Crcns: Data Sharing: a Joint Database of Experiments and Models of Reaching Movement @ Rehabilitation Institute of Chicago <a href="#">Show summary</a>	0.807	
2011 – 2015	<b>Kording, Konrad P.</b> Miller, Lee E	R01	The Representation of Uncertainty in the Sensorimotor System @ Rehabilitation Institute of Chicago <a href="#">Show summary</a>	0.892	
2012 – 2019	<b>Kording, Konrad P. (co-PI)</b> Segraves, Mark A [v]	R01	Neural Mechanisms of Fixation Choice While Searching Natural Scenes @ Northwestern University <a href="#">Show summary</a>	1	
2013	<b>Kording, Konrad P.</b>	R13	Computational and Translational Motor Control @ Rehabilitation Institute of Chicago <a href="#">Show summary</a>	0.892	

# Data: demographics

- Gender
  - Based on first name, genderize.io. Reasonable validation.
- Race
  - Based on first+last name (Author-ity 2009 dataset plus Wikipedia author profiles). Validation limited.
- Academic age
  - Inferred from PhD date
  - Used to assess seniority of mentors

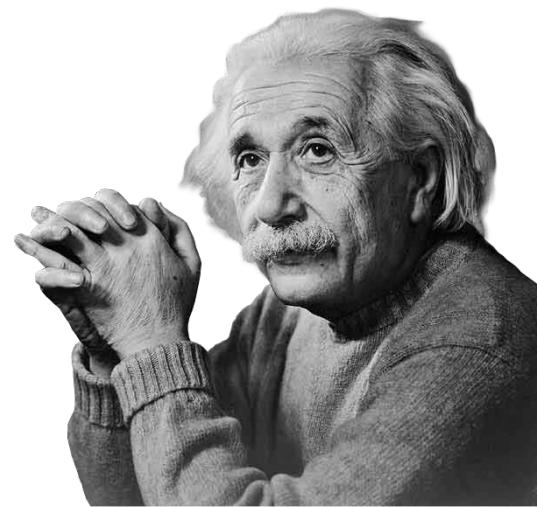
# How to measure success?

- Career success in sports
  - win MVP award for awesome performance
- Career success in academic research:
  - win Nobel prize for awesome research



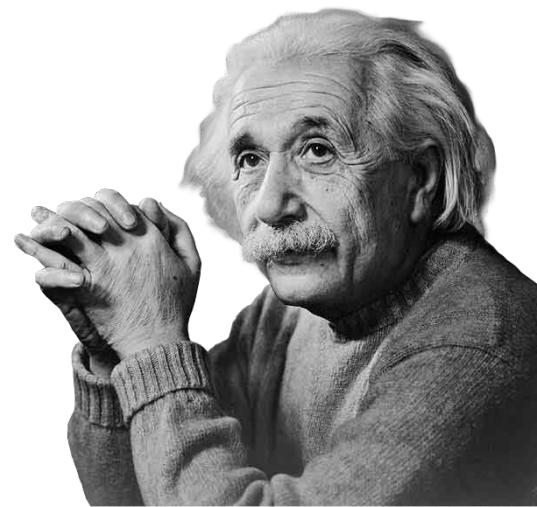
# How to measure success?

- Career success in sports
  - **play in a professional club**
  - win MVP award for awesome performance
- Career success in academic research:
  - **get a permanent faculty position**
  - win Nobel prize for awesome research



# How to measure success?

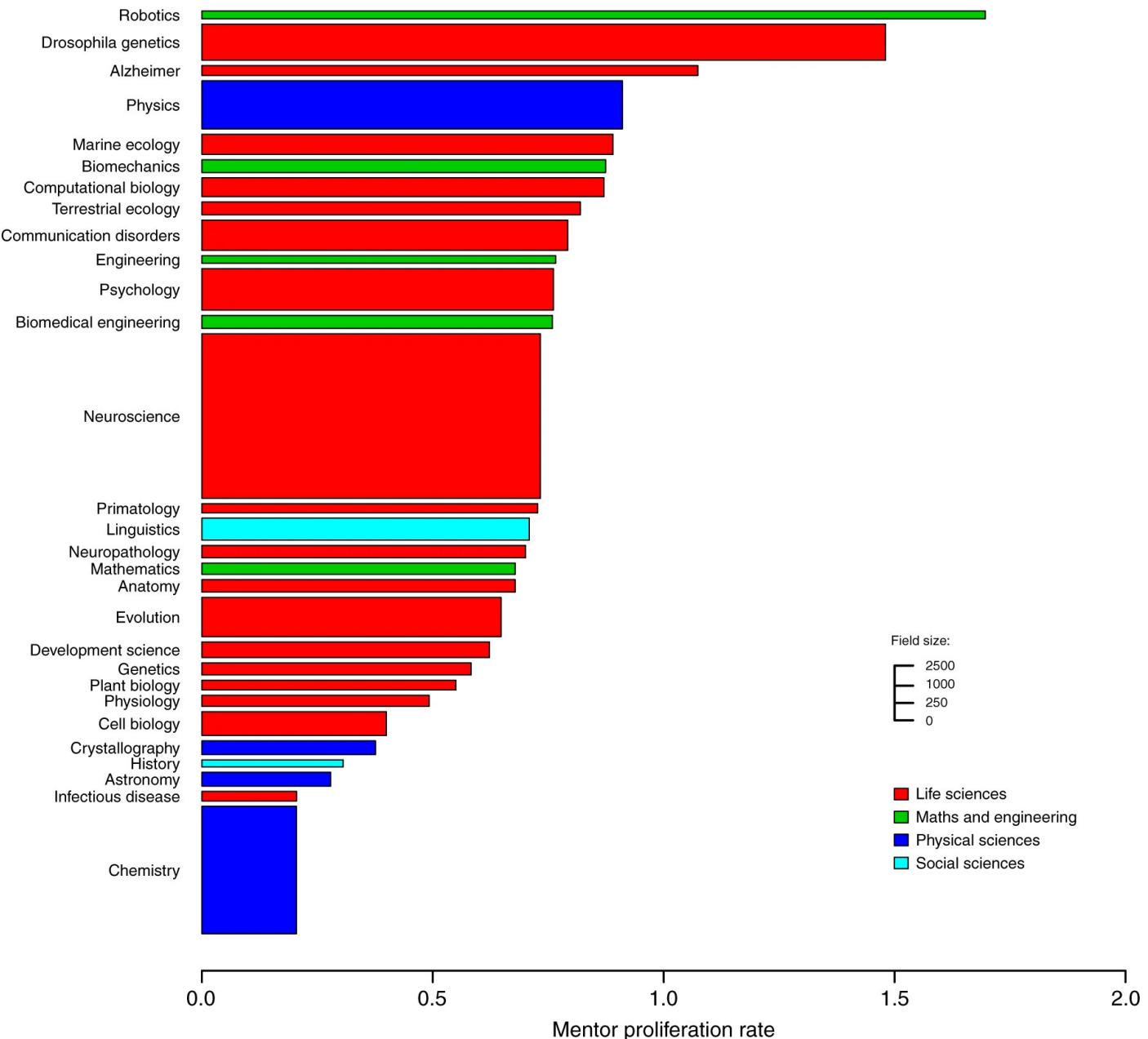
- Career success in sports
  - **play in a professional club**
  - win MVP award for awesome performance
- Career success in academic research:
  - **get a permanent faculty position**
  - win Nobel prize for awesome research
- Quantification -- number of trainees
  - 0 trainees <--> not an independent investigator
  - many trainees <--> many grants, papers, etc.



# Selection criteria

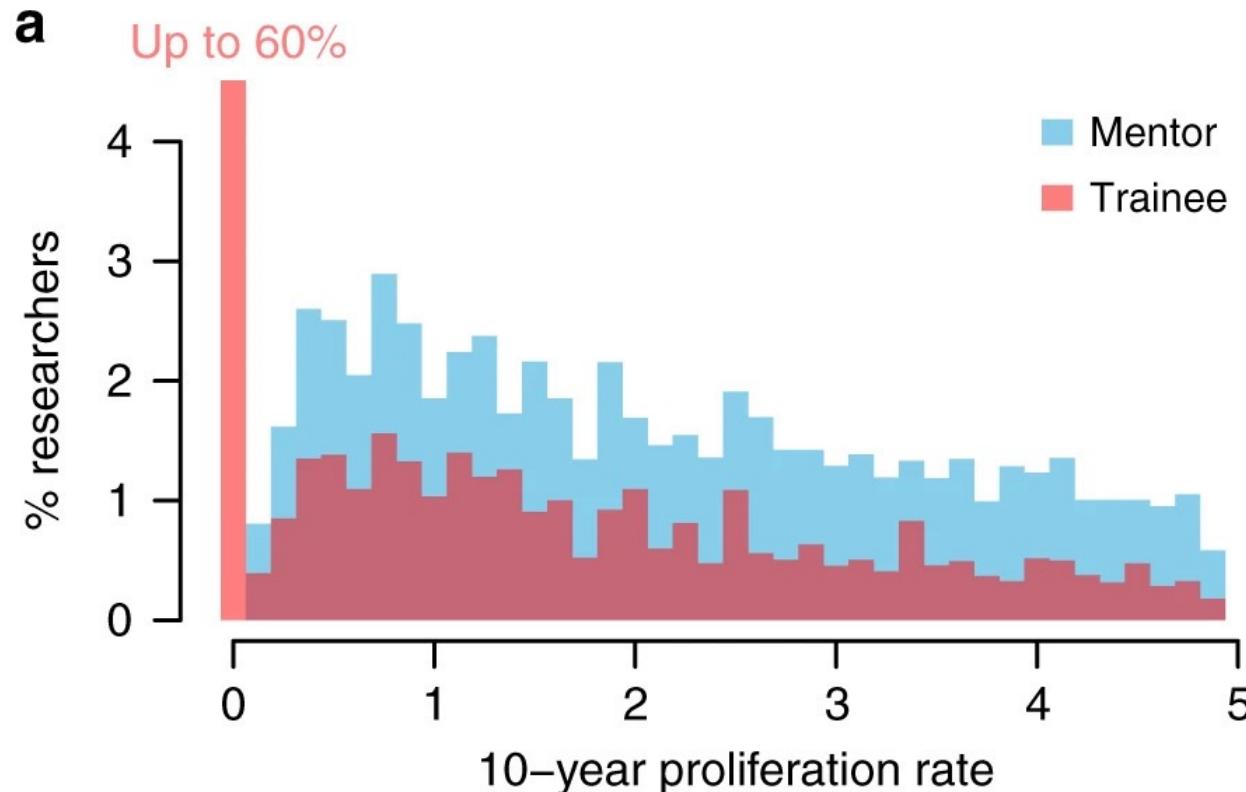
- Focus on life sciences (bulk of current data)
- ~20,000 trainees reporting...
  - graduate advisor
  - postdoc advisor
  - finished last postdoc before 2007

➤ Small, well-curated subset of the database



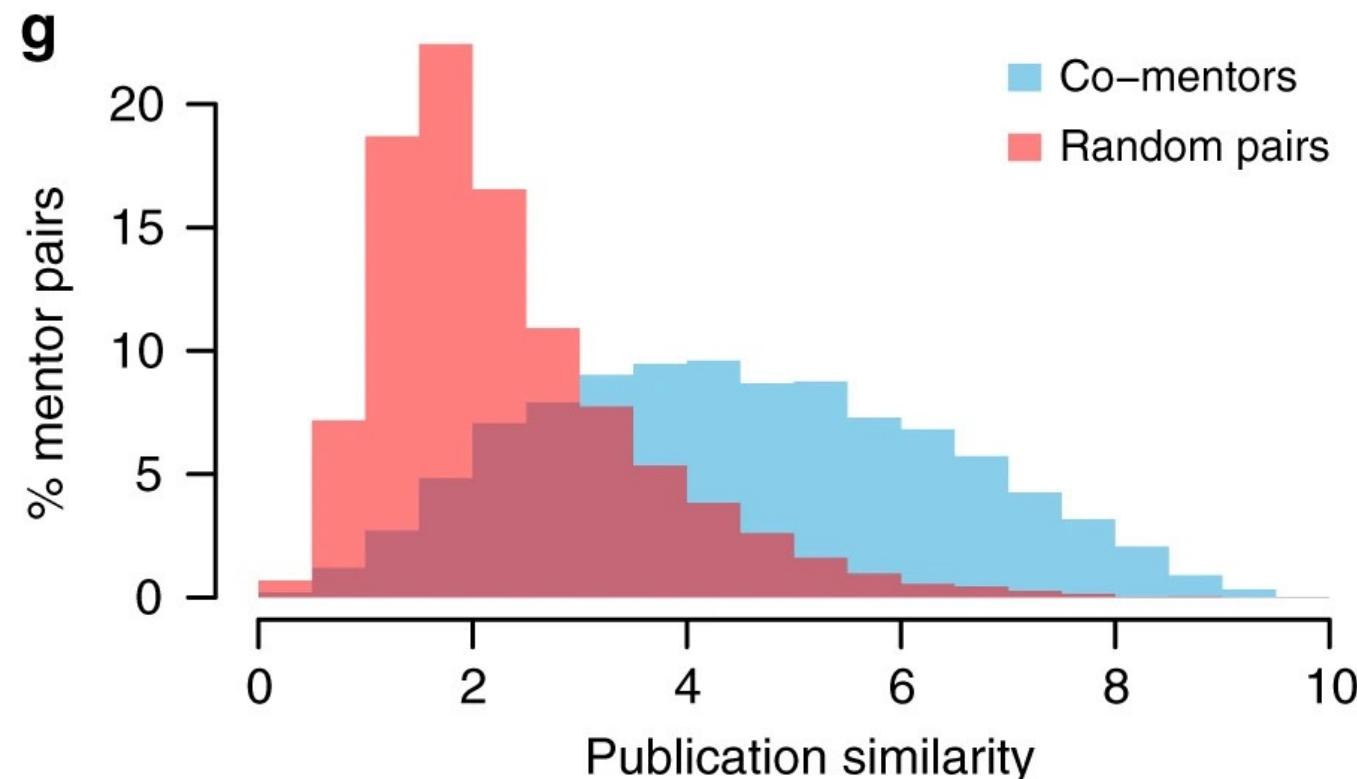
# Exploratory analysis

- Mentors and trainees similar “proliferation rates”
- Except that the majority of trainees never become mentors



# Exploratory analysis

- Co-mentors of the same trainee (their graduate and postdoc mentors) tend to work in related fields

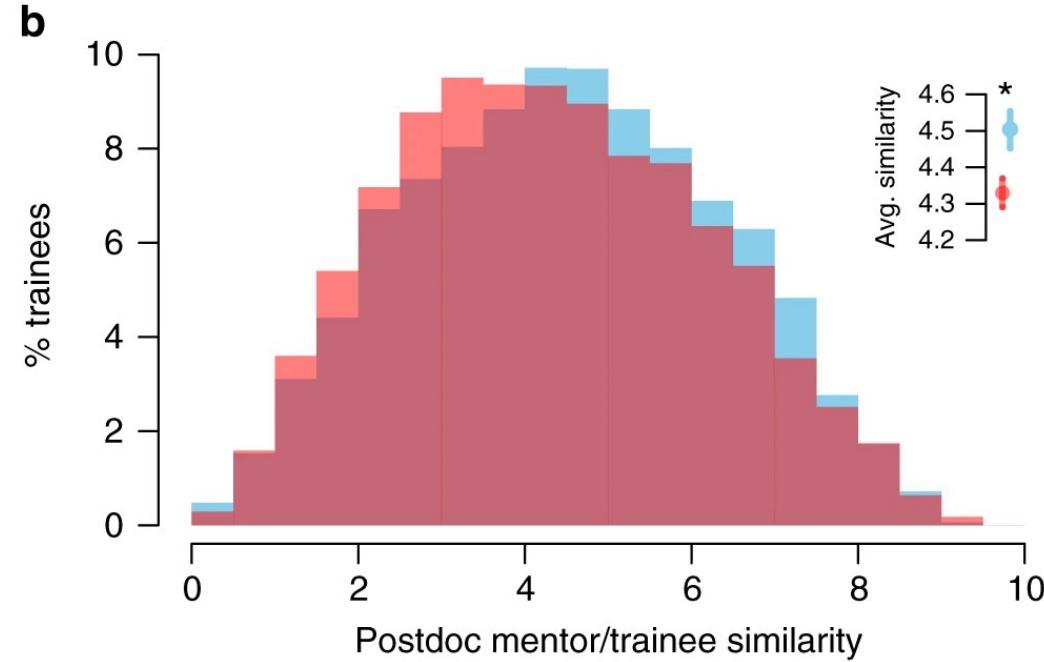
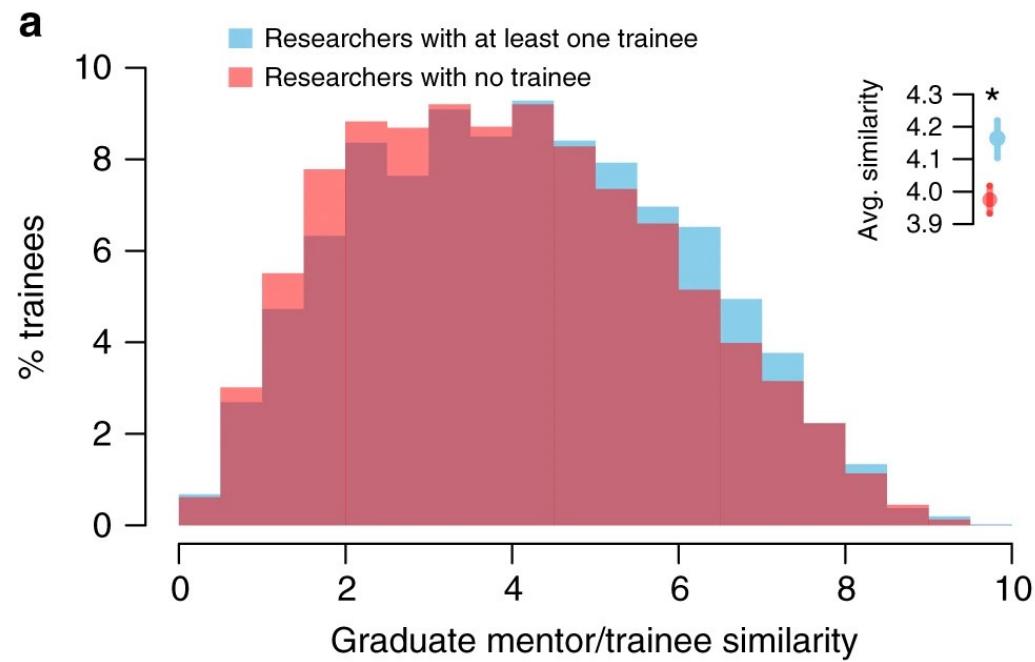


# Exploratory analysis

- Previous results are mostly confirmatory
- Effect of publication similarity on the odds of securing a permanent position?

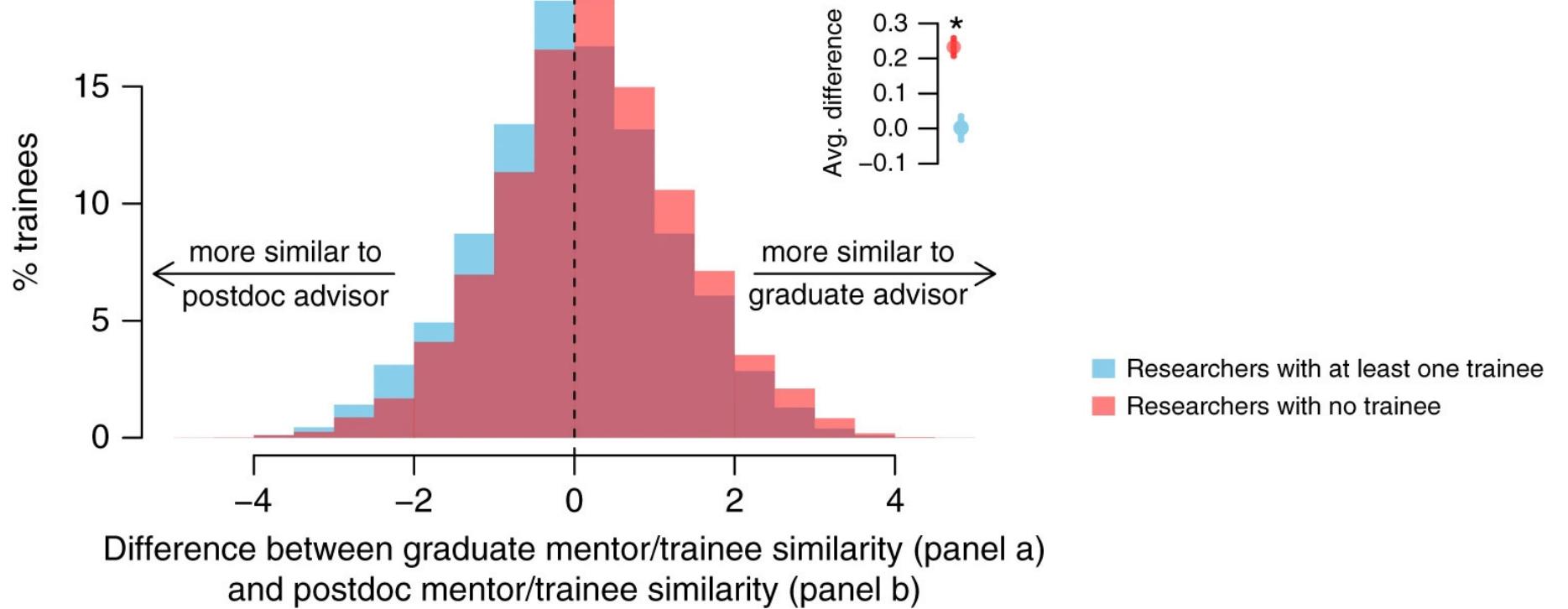
# Exploratory analysis

- Trainees that published in their mentors' topic are more successful in securing a job



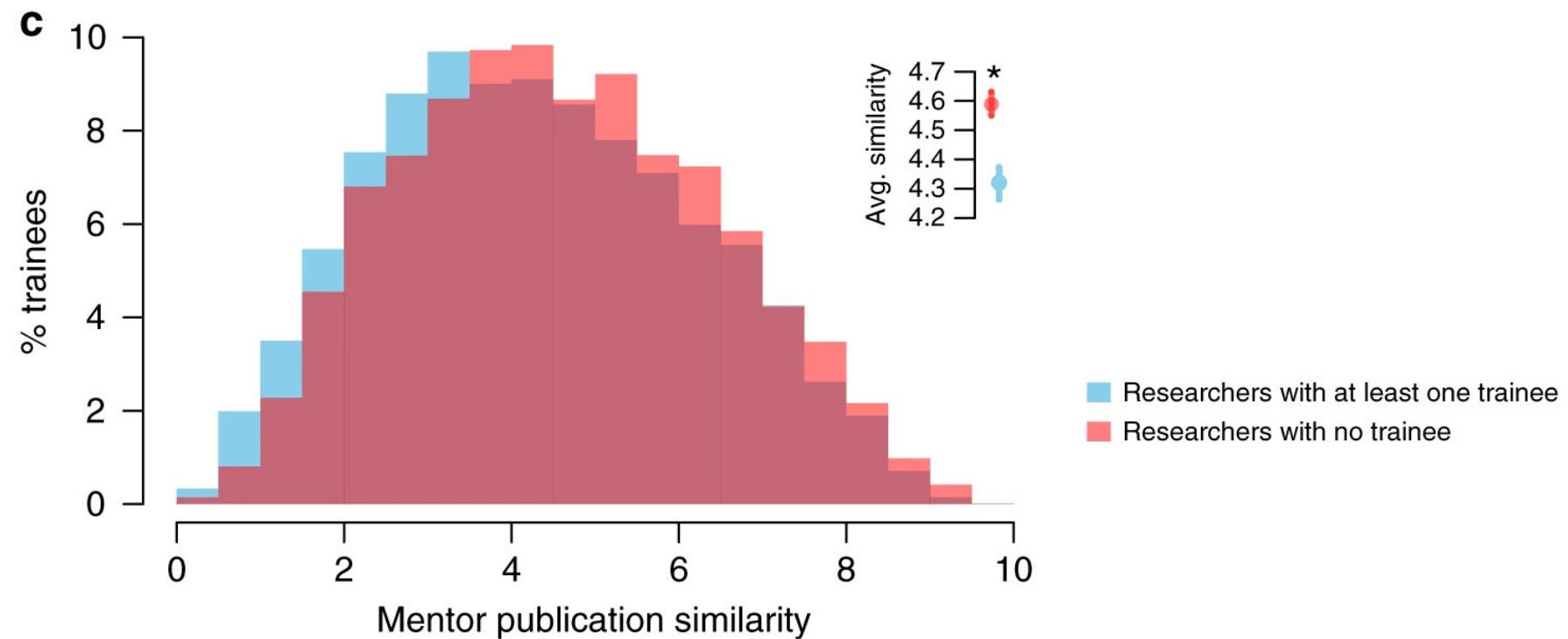
# Exploratory analysis

- Proximity to postdoc advisor is more important



# Exploratory analysis

- Different mentors have more successful trainees

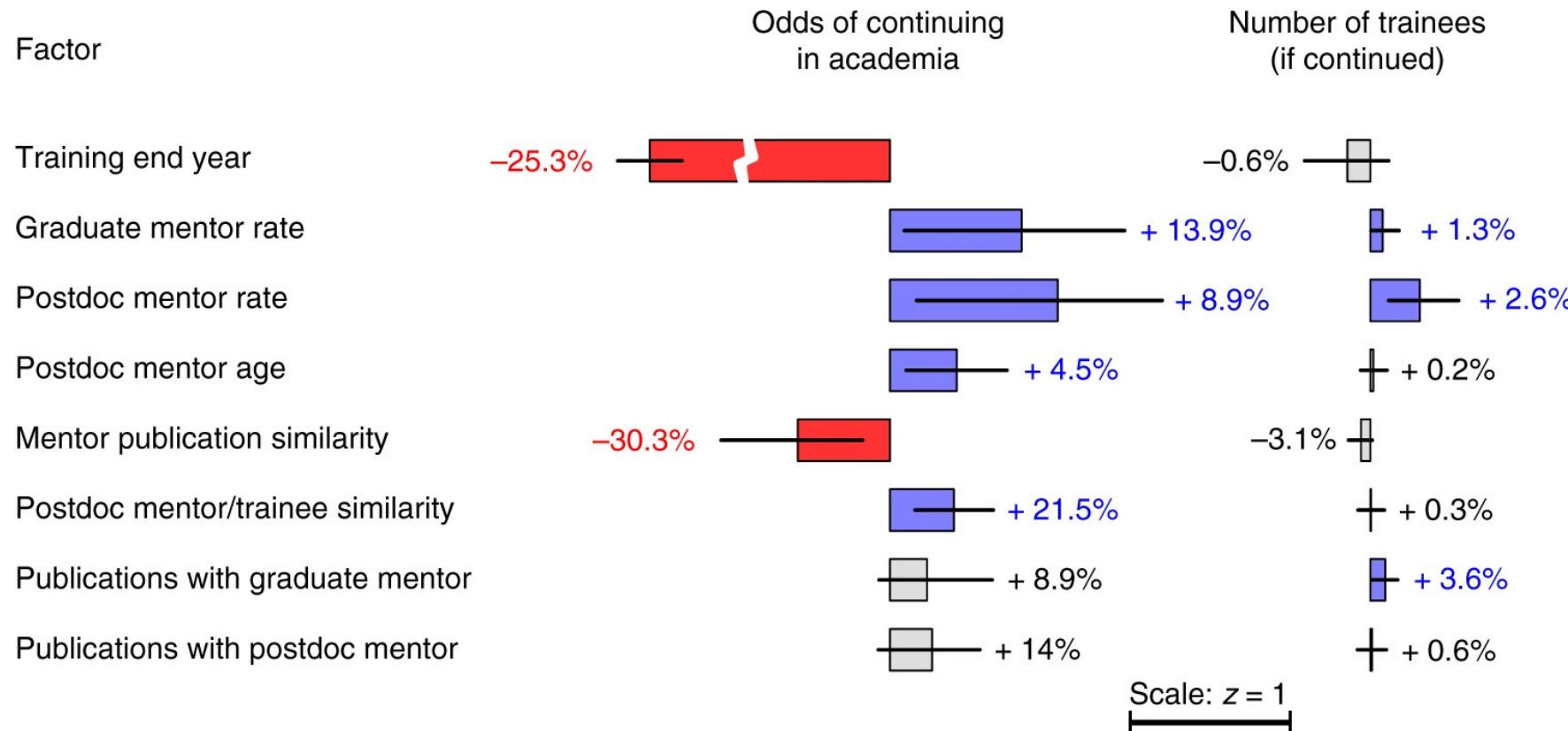


# What about confounds?

- Model! Predict “proliferation rate” from all features
  - Shapley value analysis
  - Try all combinations C of predictors
- With each C: compute the cross-validated goodness-of-fit
- Deduce the relative importance of each predictor

# Regression analysis

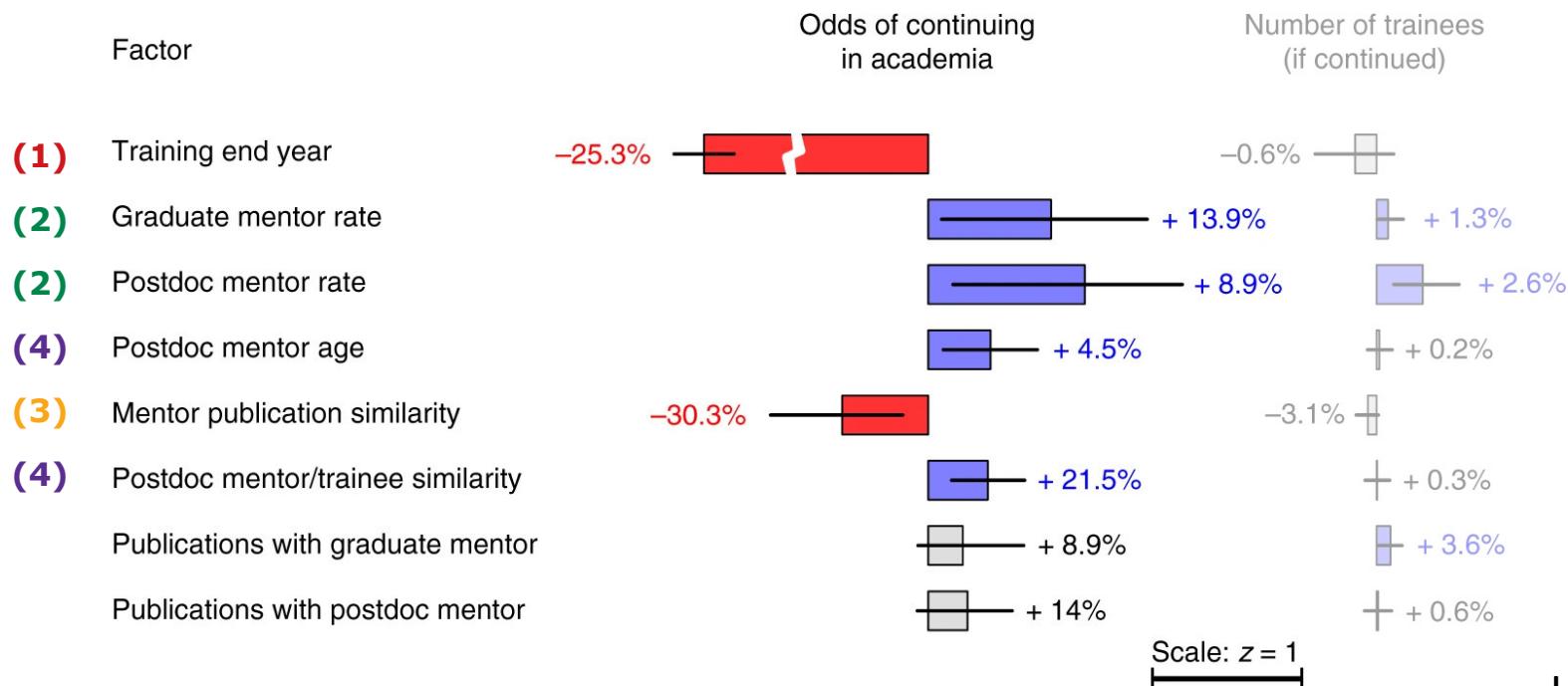
- Best model: zero-inflated negative binomial
  - Prediction continuation to independence + number of trainees



# Regression analysis

Determinants for *getting a position*:

- (1) Ph.D. before the “bottleneck years”
- (2) Successful mentors
- (3) Mentors working on different topics
- (4) Close to a senior postdoc mentor



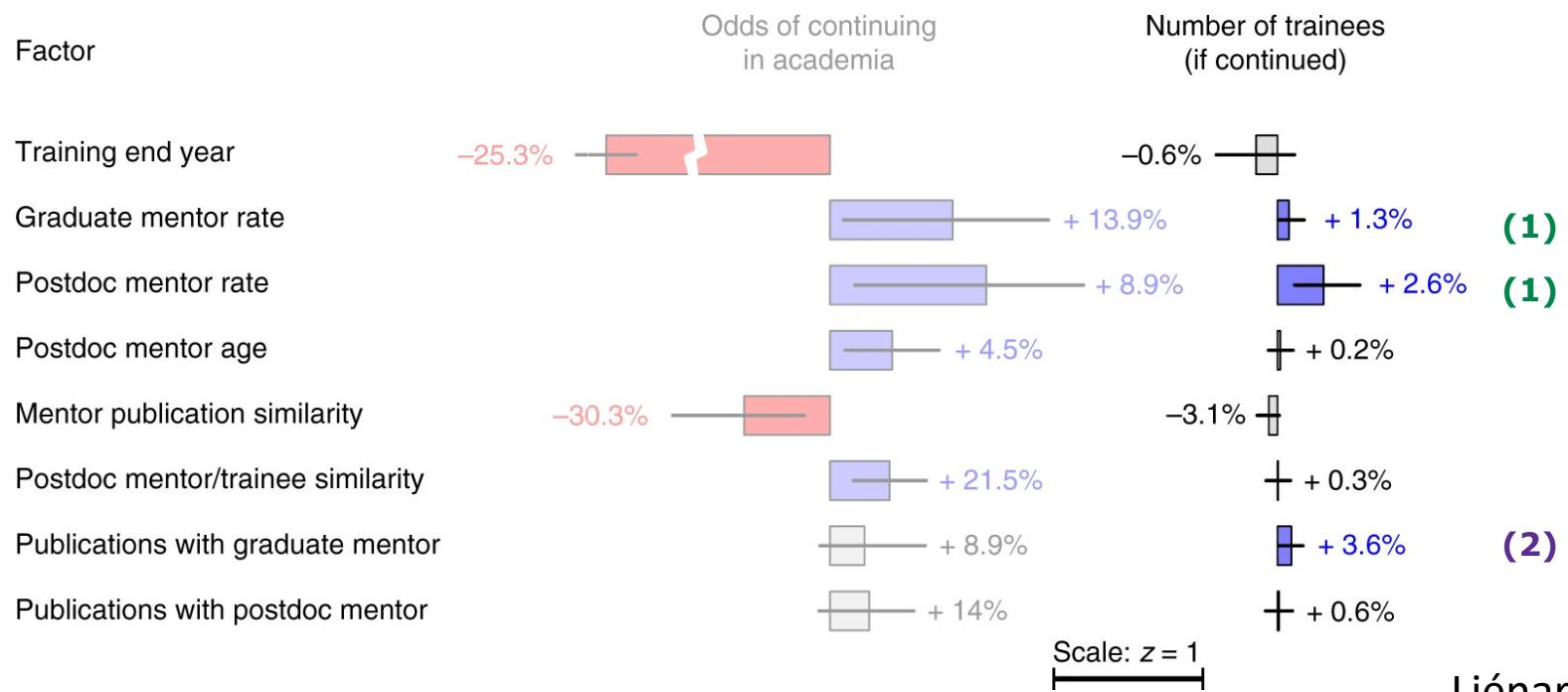
# Regression analysis

- Determinants for *having many trainees*:

(1) Successful mentors

(especially so for the postdoc mentor)

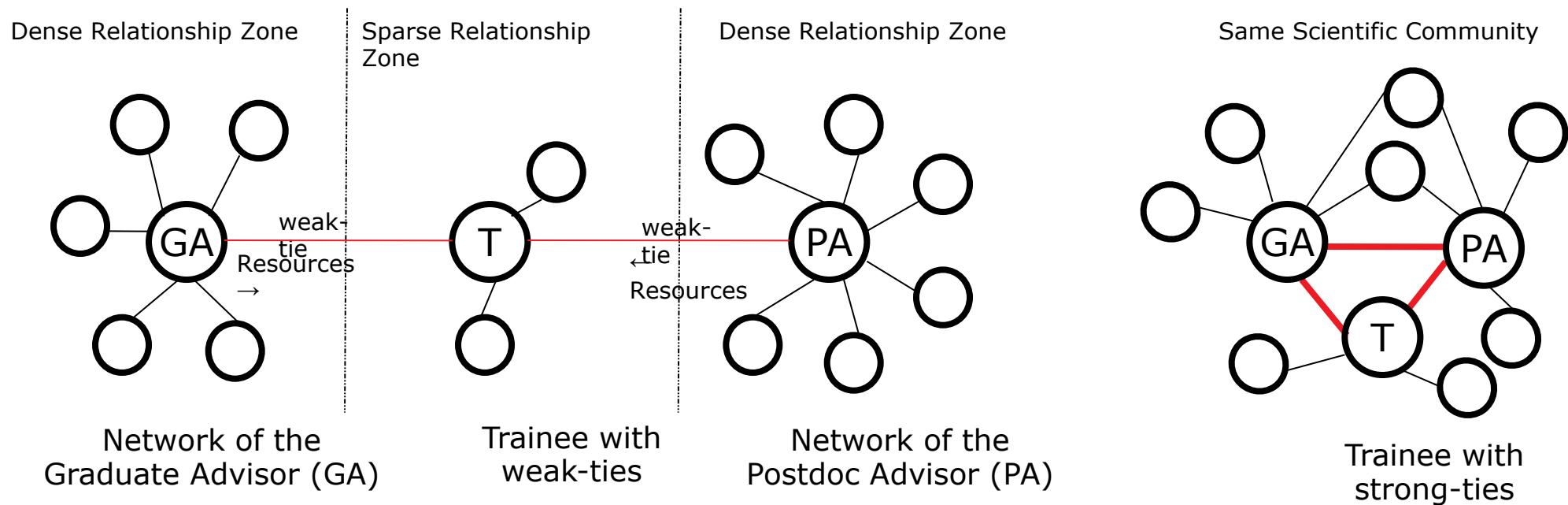
(2) Publish during Ph.D. (small effect)



# Synthesis of existing ideas/approaches is valuable

- Weak-ties theory of Granovetter (1974):

- founding dogma of much of social science
- weak ties are more useful than strong ties



# Thanks!



Jean Liénard



Titipat Achakulvisut



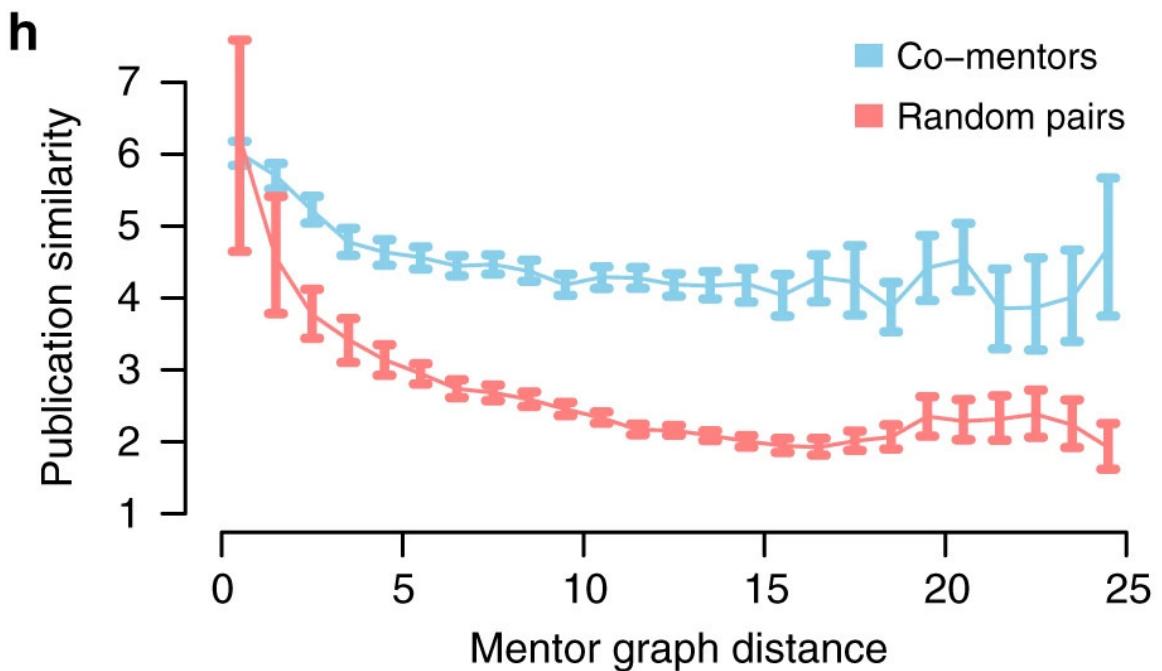
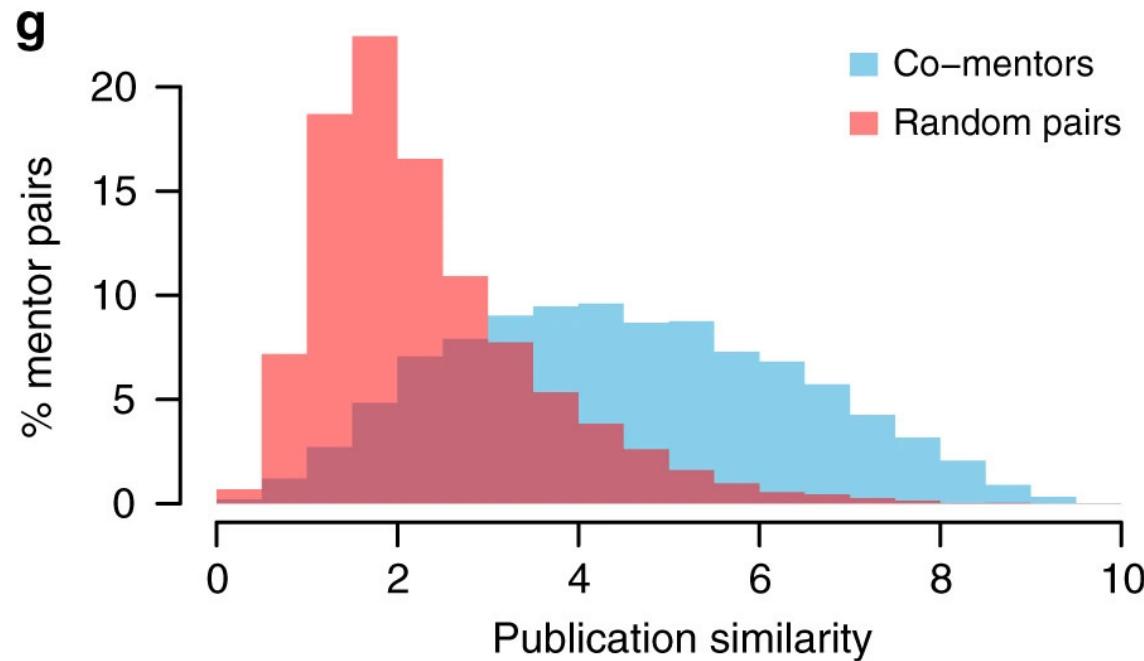
Daniel Acuna

Funding:



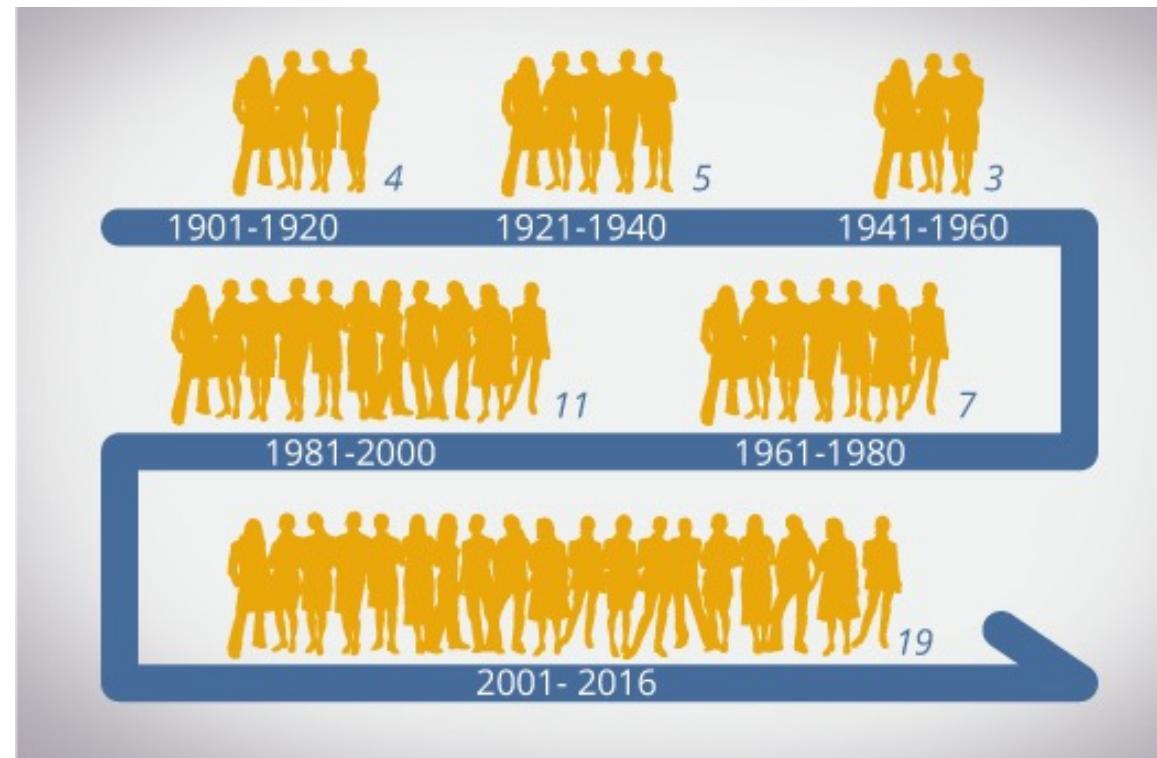
academictree.org

# To-do: Network distance vs. publication similarity?



- What is captured in the semantic similarity/dissimilarity between mentors and trainees?

# Bonus: Gender and mentorship

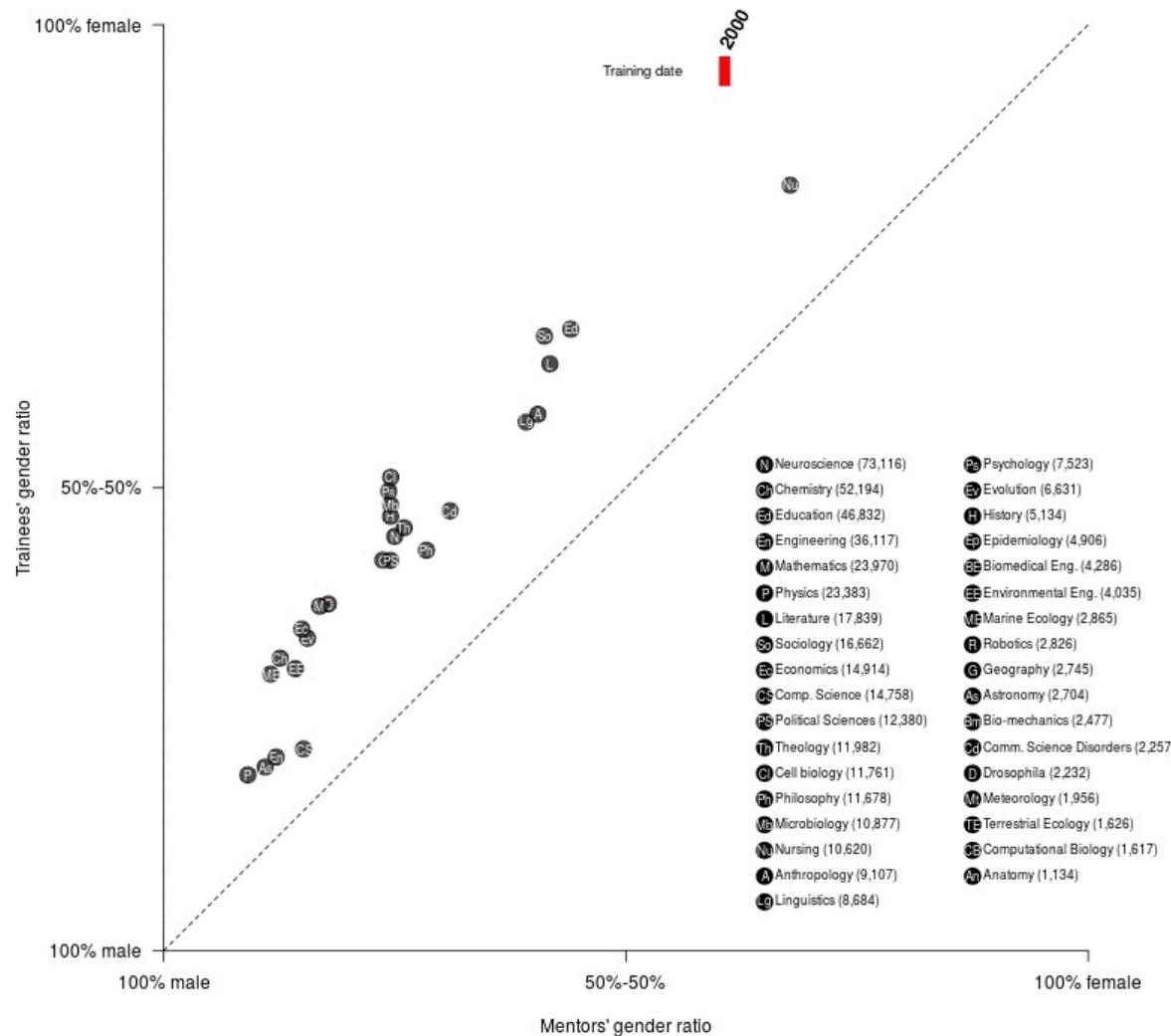


Number of women awarded Nobel Prize

*Q: how many Nobel Prizes for men in 2001-2016?*

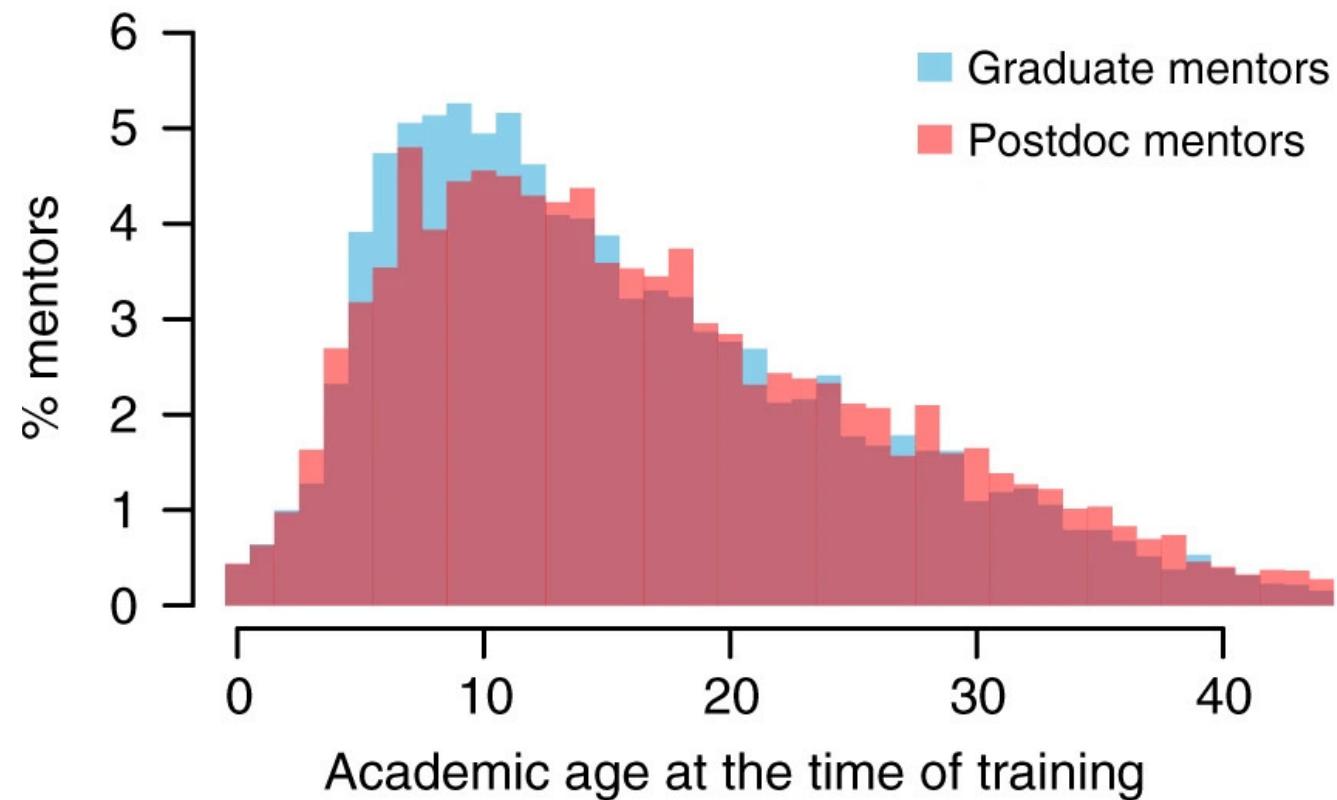
**A: 149**

# Bonus: Gender and mentorship



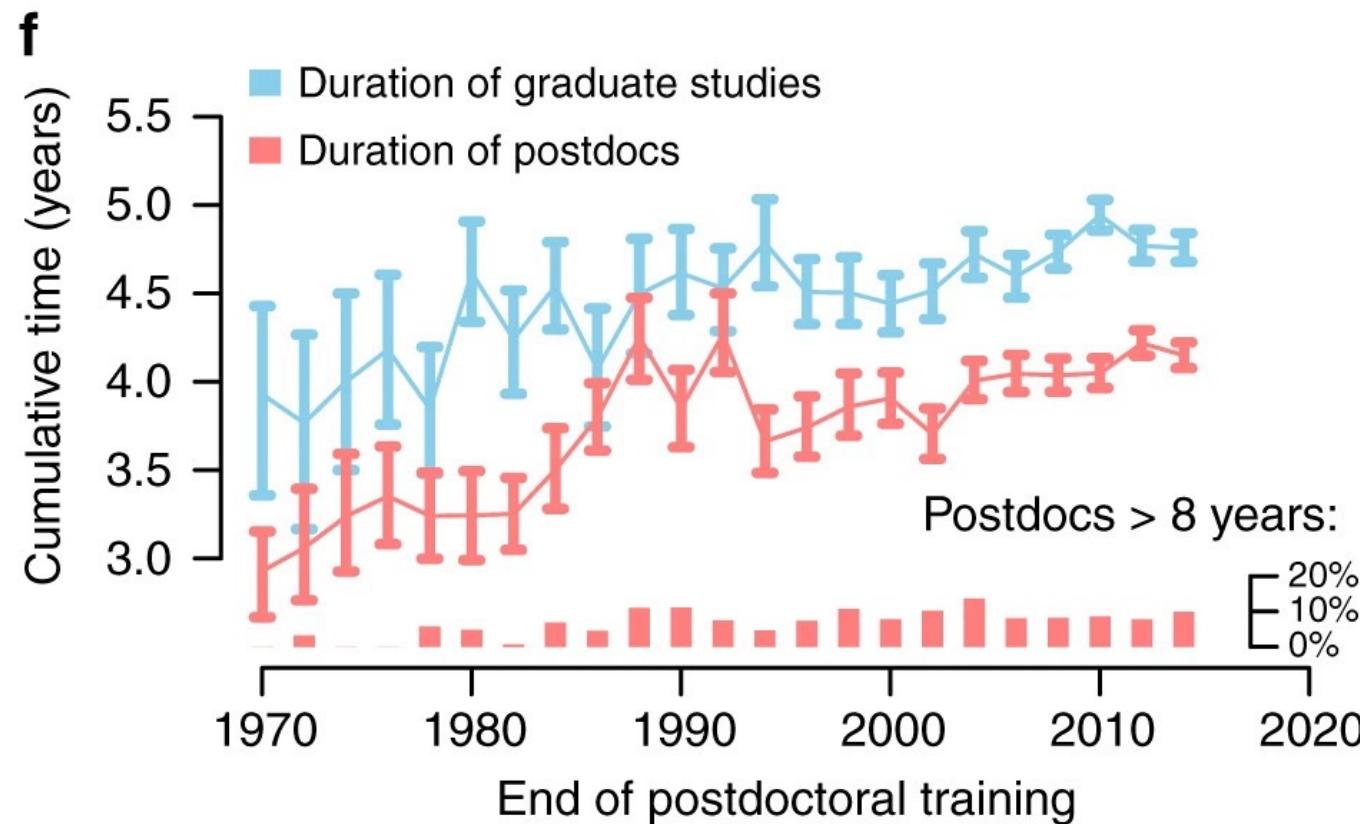
# Exploratory analysis

- “Training prime” 10 years after getting a job
- Graduate mentors are slightly more junior compared to postdoc mentors



# Exploratory analysis

- Researchers spend more and more time training
- “Permadocs” represent 10% of all postdocs since 1990s



# Data: awards

- Data sources:
  - Nobel Prize nominee
  - US National Academy of Sciences
  - Howard Hughes (prestigious in the US)
- Extracted from online listings

