Science Advances
AAAS

advances.sciencemag.org/cgi/content/full/4/8/eaat4211/DC1

Supplementary Materials for

Cross-disciplinary evolution of the genomics revolution

Alexander M. Petersen*, Dinesh Majeti, Kyeongan Kwon, Mohammed E. Ahmed, Ioannis Pavlidis*

*Corresponding author. Email: ipavlidis@uh.edu (I.P.); apetersen3@ucmerced.edu (A.M.P.)

Published 15 August 2018, *Sci. Adv.* **4**, eaat4211 (2018)
DOI: 10.1126/sciadv.aat4211

This PDF file includes:

Appendix S1. Author name disambiguation.

An important challenge in the career dataset is name disambiguation for authors and co-authors. Ambivalence primarily stems from the inconsistent use of suffixes and middle names or initials. Within the collaboration profile of $\mathcal{F}_i$, there are J_0^i raw name strings, $Name_j^i$, indexed by j. The following steps outline the disambiguation procedure we applied to address name conflicts:

A. Clean last names: Remove strings at the end of $Name_j^i$ that are not last names, and which may not be consistently listed for pollinator j across the profile of $\mathcal{F}_i$ - e.g., "Jr.", "III", and the like. At the end of this removal process, each pollinator's name string $Name_j^i$ would ideally consist of a first name string FN_j^i, possibly a middle name string MN_j^i, and a last name string LN_j^i.

B. Disambiguate middle initial strings within each $\mathcal{F}$ profile: Within the profile of each $\mathcal{F}_i$, search for inconsistencies in the use of $MN_{j?}^i$. For example, sometimes pollinator j? may be listed as *Amanda M Price*, some other times as *Amanda Price*, and yet some other times as *Amanda Miranda Price*. In this example, the last name string $LN_{j?} = Price$ and the first name string $FN_{j?} = Amanda$ are consistent. However, the middle name string set $\{\_, M, Miranda\}$ introduces ambiguity, as it includes instances of no middle name $MN_{j?}^i \equiv \emptyset$, middle initial $MN_{j?}^i \equiv \mathcal{I}$, and full middle name $MN_{j?}^i \equiv \mathcal{I}name$, compatible with the middle initial instances. Depending on the type of middle name ambiguity for pollinator j?, apply the following rules:

- If the middle name ambiguity set for pollinator j? has instances of no middle name and middle initial only $\{\emptyset, \mathcal{I}\}$, then transform all name instances of pollinator j? to $\langle FN_j^i \quad \mathcal{I} \quad LN_j^i \rangle$, firmly assigning them the j index.

- If the middle name ambiguity set for pollinator j? has instances of no middle name, middle initial, and full middle name, compatible with the middle initial instances $\{\emptyset, \mathcal{I}, \mathcal{I}name\}$, then transform all name instances of pollinator j? to $\langle FN_j^i \quad \mathcal{I}name \quad LN_j^i \rangle$, firmly assigning them the j index.

- If the middle name ambiguity set for pollinator j? has instances of no middle name and two different middle initials $\{\emptyset, \mathcal{I}1, \mathcal{I}2\}$, then check if $\langle FN_{j?}^i \quad LN_{j?}^i \rangle$ and $\langle FN_{j?}^i \quad \mathcal{I}1 \quad LN_{j?}^i \rangle$ are co-authors in the same paper within the $\mathcal{F}$ profile i. If they are, then transform $\langle FN_{j?}^i \quad LN_{j?}^i \rangle$ to $\langle FN_{j_2}^i \quad \mathcal{I}2 \quad LN_{j_2}^i \rangle$, assigning to the consolidated subset a j_2 index.
  If the first test fails, then check if $\langle FN_{j?}^i \quad LN_{j?}^i \rangle$ and $\langle FN_{j?}^i \quad \mathcal{I}2 \quad LN_{j?}^i \rangle$ are co-authors in the same paper within the $\mathcal{F}$ profile i. If they are, then transform $\langle FN_{j?}^i \quad LN_{j?}^i \rangle$ to $\langle FN_{j_1}^i \quad \mathcal{I}1 \quad LN_{j_1}^i \rangle$, assigning to the consolidated subset a j_1 index.
  If both tests fail, then compare the co-authors among $\langle FN_{j?}^i \quad LN_{j?}^i \rangle$, $\langle FN_{j?}^i \quad \mathcal{I}1 \quad LN_{j?}^i \rangle$, and $\langle FN_{j?}^i \quad \mathcal{I}2 \quad LN_{j?}^i \rangle$ within the $\mathcal{F}$ profile i. Transform the no middle name instances to the middle name variety with which it shares more co-authors.

C. Disambiguate pollinators across $\mathcal{F}$ profiles: Let j and j' be pollinators in $\mathcal{F}$ profiles i and i', respectively. Check if j and j' are likely the same person, $j \equiv j'$, in order to establish (or not) a mediated association link between i and i'. Depending on the type of ambiguity, apply the following rules:

- If the ambiguous pollinators have the same first names, FN_j^i=$FN_{j'}^{i'}$, and the same last names, $LN_j^i = LN_{j'}^{i'}$, and the middle name ambiguity set consists of no middle name plus at least two different middle names $\{\emptyset, \mathcal{I}name1, \mathcal{I}name2, \ldots\}$, then compare the common co-authors among the no middle name instances and the other instances, assigning the no middle name instance to the case with which it shares the most common co-authors.

- If the first name of a pollinator j is hyphenated $FN_j^i \equiv FN1_j^i - FN2_j^i$, and has only 2 letters i.e., $FN1$ and $FN2$ are one letter, check for any other pollinator j' that has hyphenated first name with the same first letter $FN1_j^i$ and the first letter after the hyphen starts with $FN2_j^i$. Then transform the j pollinator to j' who has the longest such hyphenated first name.
- If the first name of a pollinator j is hyphenated $FN_j^i \equiv FN1_j^i - FN2_j^i$, check for any other pollinator j' that has $FN_{j'}^{i'} = FN1_j^i$, $\mathcal{J}' = FN2_j^i$, and $LN_{j'}^{i'} = LN_j^i$. If such a pollinator j' does exist and shares at least one common co-author with j, then transform the j pollinator to j', assigning the second part of her/his original hyphenated first name to be her/his middle name.
- If the name of a pollinator j has only two letters $FN_j^i \equiv \mathcal{L}1_j^i - \mathcal{L}2_j^i$, check for any other pollinator j' that has $FN_{j'}^{i'} = \mathcal{L}1name_j^i$, $\mathcal{J}' = \mathcal{L}2name_j^i$, and $LN_{j'}^{i'} = LN_j^i$. If such a pollinator j' does exist and shares at least one common co-author with j, then transform the j pollinator to j'.

Appendix S2. Connectivity of the $\mathcal{F}$ network.

How does the $\mathcal{F}$ network depend on the direct $\mathcal{F}$ connectivity? To investigate this, we randomly removed a fraction q of the links, incrementing q over the range $[0,1]$, and monitoring the effect on the network's giant and non-giant components. This method of random link removal is drawn from the theory of phase transitions in the connectivity of networks (62,63). For each q, we performed the link percolation 40 times and reported the mean and standard deviation of the following network connectivity descriptors:

Giant component size:
For the $\mathcal{F}$ collaboration network, the initial size of the largest connected component (aka giant component) is $S_G(q = 0) = 3{,}869$, meaning that 321 $\mathcal{F}$ nodes are initially disconnected from the giant component. Figure S1a shows the ratio $S_G(q)/S_G(q = 0)$ as a function of q, demonstrating the robustness of the collaboration network - even after 80% of the links are removed, roughly 60% of the $\mathcal{F}$ are still connected within the network. Of course, the fragmentation of the network depends on how we remove the links. We compared the results for random uniform removal of links and for random removal according to the weight $W_{ii'}$. For each $W_{ii'}$ definition, we removed the links according to increasing weight and also according to the inverted weight $W_{ii'}^r = \max_{ii'}[W_{ii'}] - W_{ii'}$ ('reverse'). We used three definitions for the link weights $W_{ii'}$: (a) $W_{ii'} \equiv \max[PR_i, PR_{i'}]$, where PR_i and $PR_{i'}$ are the PageRank centralities of node i and i', respectively, using the common damping factor 0.85; (b) $W_{ii'} \equiv \max[B_i, B_{i'}]$, where B_i and $B_{i'}$ are the betweenness centralities of nodes i and i', representing the number of shortest paths in the network that traverse i and i', respectively; (c) $W_{ii'} \equiv O_{ii'}$, where $O_{ii'} \in [0,1]$ is the overlap fraction in the first-degree neighbors of nodes i and i', calculated as $O_{ii'} = s_{ii'}/[(k_i - 1) + (k_{i'} - 1) - s_{ii'}]$, where $s_{ii'}$ is the number of shared first-degree neighbors, and k_i and $k_{i'}$ are the degrees of nodes i and i', respectively (64). Consistent with expectations, the link removal methods that exhibited the sharpest fragmentation were $PR_{ii'}$ and $B_{ii'}$.

Susceptibility to fragmentation:
For each q we calculated the size S_i of all the N_q fragments, where by definition $S_G(q) = \max_i(S_i(q))$. The severity of the fragmentation (percolation) process can be further illustrated by analyzing the fragment size distribution $P(S_i)$, i.e., by calculating the distribution's second moment $\sigma_q^2 = \sum_{i|S_i<S_G}^{N_q-1} S_i^2\, P(S_i)$. By construction, σ_q^2 does not include the giant component S_G. The fluctuation scale σ_q^2 diverges when the network shatters into pieces of varying sizes. Indeed, fig. S1b shows how the

network's susceptibility to fragmentation peaks - depending on the link removal weights – when there is a precipitous drop in the connectivity of the giant component (fig. S1a). The fragmentation peak is associated with the critical point of the network, and is achieved at a smaller q value when the links associated with the most central $\mathcal{F}$ are removed first (blue and black curves in fig. S1b).

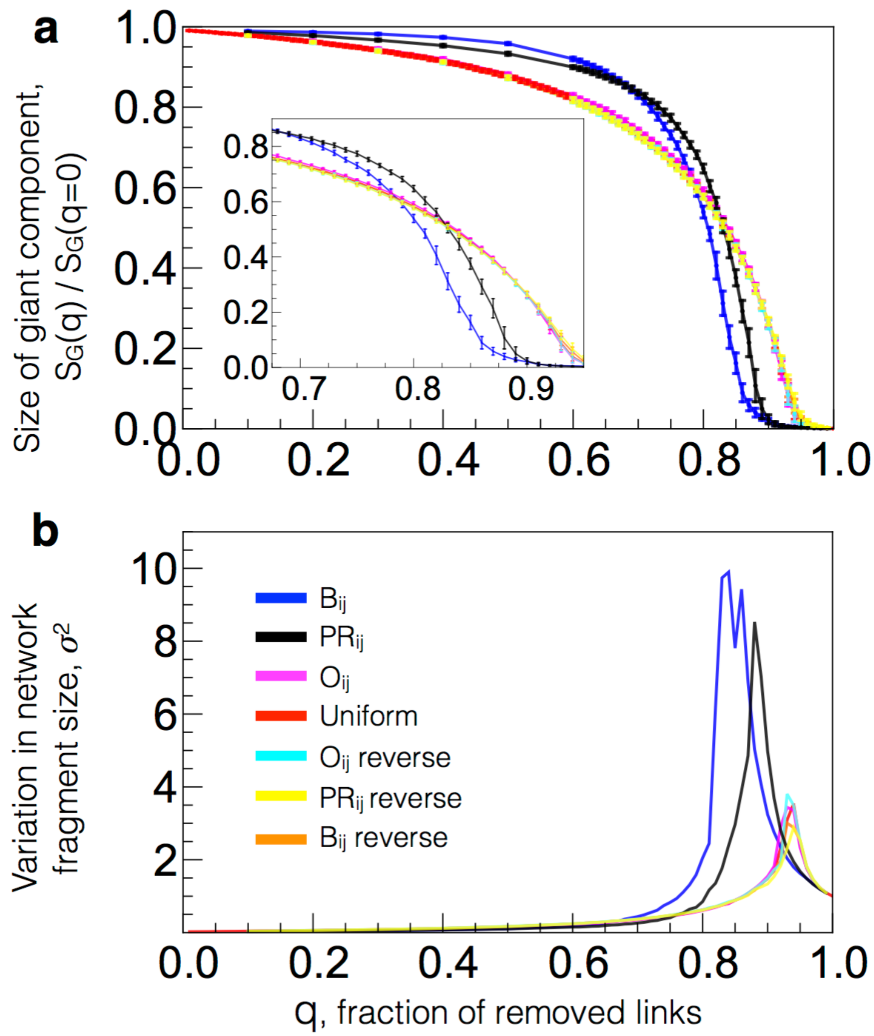

Fig. S1. Robustness of the $\mathcal{F}$ network with respect to link removal. (a) The ratio $S_G(q)/S_G(q=0)$ measures the size of the largest remaining fragment $S_G(q)$, relative to the size of the initial giant component $S_G(q=0)$. The slow decay until $q=0.6$ indicates that this network is robust to variation in the connectivity of scholars. For a given q, we repeated the fragmentation process 40 times, and plotted the error bars to indicate the mean and standard deviation. (b) Detection of the critical point at which the college disassociates. For each q we also monitor the size S_i of all the N_q disconnected network fragments, where by definition $S_G(q) = \max_i(S_i(q))$. As a limiting example, complete disassociation occurs for $q=1$ (all links removed), corresponding to a completely disconnected ensemble of nodes with $N_q = 4{,}190$ and $S_i = 1$ for all i. The fluctuation scale of the fragmentation process is illustrated by the variation of the fragment size distribution, σ_q^2, which diverges when the network 'shatters' into pieces of highly variable sizes. The peak in σ_q^2 signals the onset of the shattering process.

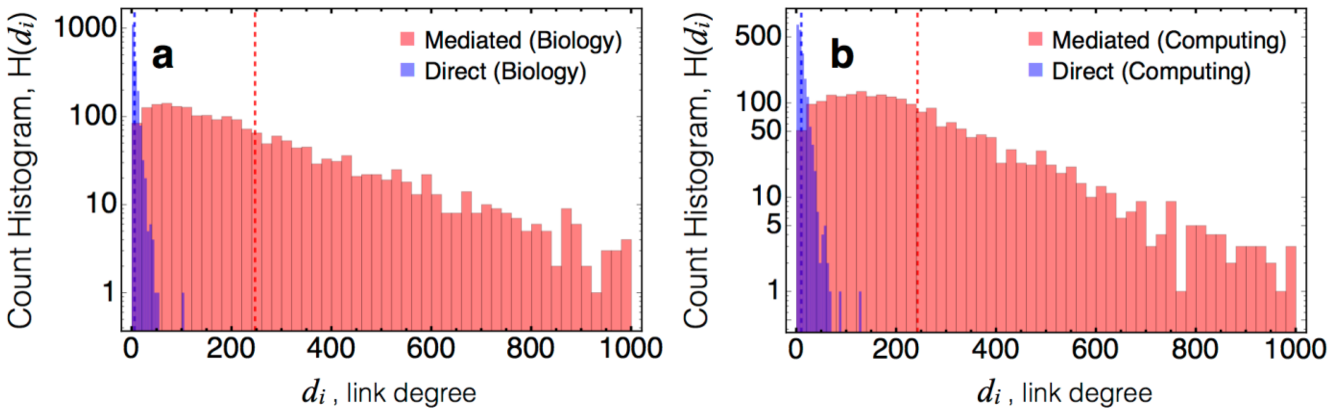

Fig. S2. $\mathcal{F}$ network distributions for direct and mediated associations. for a. biology and b. computing. Each panel shows the frequency distribution (counts) of faculty $\mathcal{F}$ with a given link degree counting the number of links for a given node, $d_i \equiv \mathscr{C}_i^D(t)$, within a particular definition of the $\mathcal{F}$ network (vertical lines indicate distribution means). The direct subnetworks only include direct links, which are established whenever two $\mathcal{F}$ collaborate on at least one publication. The mediated subnetworks only include indirect links between two $\mathcal{F}$ who have both collaborated with a common pollinator (i.e., are associated via triadic closure - see Fig. 1). On average, 97% in biology and 92% in computing are *pollinator* co-authors, i.e., researchers not included in the $\mathcal{F}$ set. The significantly different scale of the degree distributions demonstrates the connectivity power of the pollinators within the invisible college.

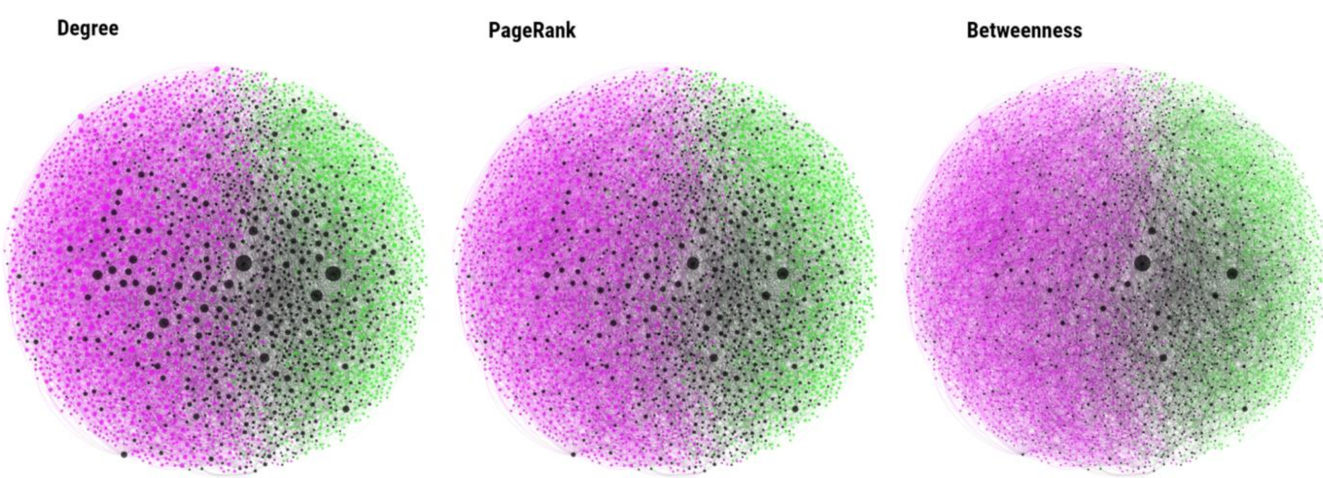

Fig. S3. Three perspectives on the centrality of $\mathcal{F}_i$ in the direct collaboration network. Shown is the giant connected component of the faculty network $\mathcal{F}$ using all data up to 2015. The nodes and links across each network are fixed, only the node sizes vary according to the indicated centrality measure: (a) degree $\mathscr{C}_i^D$, (b) PageRank $\mathscr{C}_i^{PR}$, (c) betweenness $\mathscr{C}_i^B$. Notably, the most central $\mathcal{F}_i$ according to each of the three measures is Eric Lander, one of the leaders of the HGP.

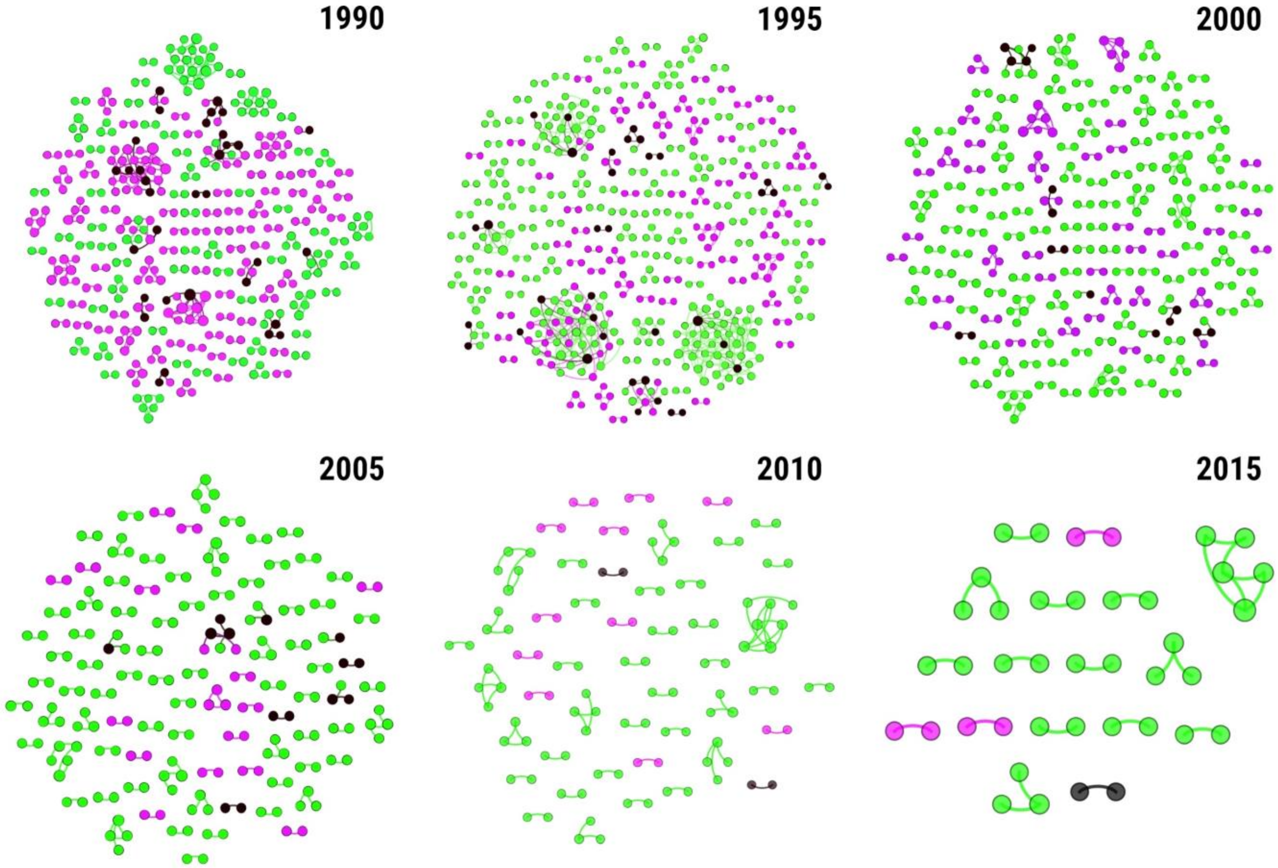

Fig. S4. Evolution of the nongiant components in the $\mathcal{F}$ network. Green and magenta nodes represent faculty $\mathcal{F}_i$ with $BIO_{\mathcal{F}}$ and $CS_{\mathcal{F}}$ affiliation, respectively; black nodes represent faculty $\mathcal{F}_i$ that by time t collaborated with at least one faculty from the opposite department and thus joined the $XD_{\mathcal{F}}$ group.

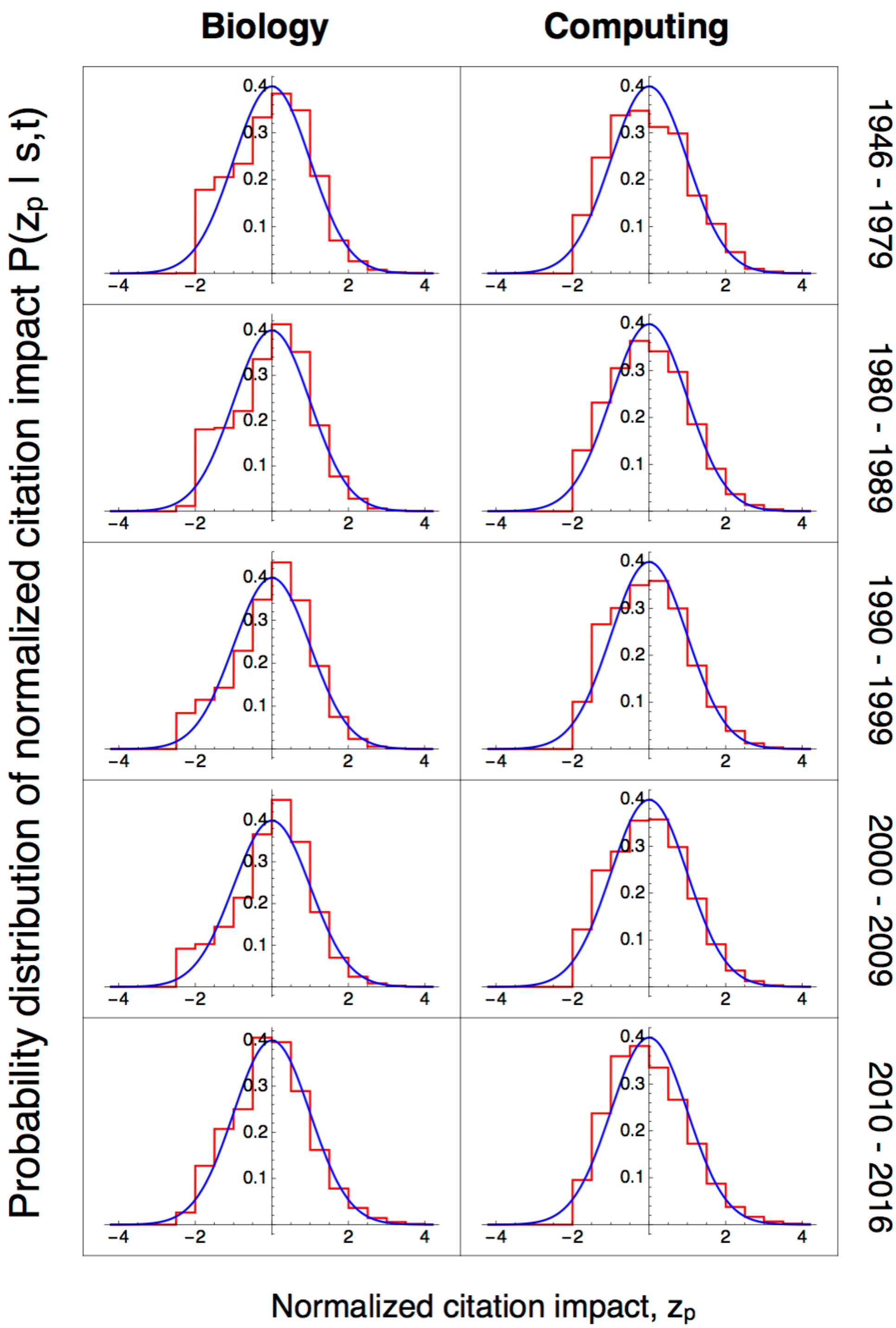

Fig. S5. Distribution of normalized citation impact by departmental affiliation and time period. Probability distribution $P(z|s,t)$ calculated by separating the publications of the career dataset into subsets according to the departmental affiliation of $\mathcal{F}_i$ and publication year. Shown are the empirical distribution (red bins) and baseline normal distribution $N(0,1)$ (blue curve), which demonstrates the time-independence of the normalized citation impact variable.

Table S1. Set of 155 biology and computing departments in the United States. for the formation of the career dataset. Ranks are per the 2014 U.S. News & World Report.

Rank	Biology Departments	Rank	Computing Departments
1	Harvard University	1	Carnegie Mellon University
1	Massachusetts Institute of Technology	1	Massachusetts Institute of Technology
1	Stanford University	1	Stanford University
4	University of California, Berkeley	1	University of California, Berkeley
5	California Institute of Technology	5	University of Illinois, Urbana Champaign
5	Johns Hopkins University	6	Cornell University
7	University of California, San Francisco	6	University of Washington
7	Yale University	8	Princeton University
9	Princeton University	9	Georgia Institute of Technology
9	Scripps Research Institute	9	University of Texas, Austin
11	Cornell University	11	California Institute of Technology
11	Duke University	11	University of Wisconsin, Madison
11	Washington University in St. Louis	13	University of California, Los Angeles
14	Columbia University	13	University of Michigan, Ann Arbor
14	Rockefeller University	15	Columbia University
14	University of California, San Diego	15	University of California, San Diego
14	University of Chicago	15	University of Maryland, College Park
18	University of Wisconsin, Madison	18	Harvard University
19	University of California, Davis	19	University of Pennsylvania
19	University of California, Los Angeles	20	Brown University
19	University of Michigan, Ann Arbor	20	Purdue University, West Lafayette
19	University of Pennsylvania	20	Rice University
19	University of Texas Southwestern Medical Center	20	University of Southern California
19	University of Washington	20	Yale University
25	Baylor College of Medicine	25	Duke University
26	Cornell University (Weill)	25	University of Massachusetts, Amherst
26	Northwestern University	25	University of North Carolina, Chapel Hill
26	University of North Carolina, Chapel Hill	28	Johns Hopkins University
26	Vanderbilt University	29	New York University
30	Emory University	29	Pennsylvania State University, University Park
30	University of Colorado, Boulder	29	University of California, Irvine
30	University of Illinois, Urbana Champaign	29	University of Minnesota, Twin Cities
30	University of Texas, Austin	29	University of Virginia
34	Brown University	34	Northwestern University
34	Indiana University, Bloomington	34	Ohio State University
34	University of California, Irvine	34	Rutgers, The State University of New Jersey
34	University of Minnesota, Twin Cities	34	University of California, Davis
38	Case Western Reserve University	34	University of California, Santa Barbara
38	Dartmouth College	34	University of Chicago
38	Mayo Medical School	40	Dartmouth College
38	University of Arizona	40	Stony Brook University, SUNY
42	Carnegie Mellon University	40	Texas A&M University, College Station
42	Icahn School of Medicine at Mount Sinai	40	University of Arizona
42	Ohio State University	40	University of Colorado, Boulder
42	Pennsylvania State University, University Park	40	University of Utah
42	Rice University	40	Virginia Tech
42	University of Alabama, Birmingham	40	Washington University in St. Louis
42	University of Georgia	48	Arizona State University
42	University of Pittsburgh	48	Boston University
50	Michigan State University	48	North Carolina State University
50	University of California, Santa Barbara	48	University of Florida
50	University of Massachusetts Medical Center	52	Indiana University, Bloomington
50	University of Virginia	52	Rensselaer Polytechnic Institute
50	Yeshiva University (Einstein)	52	University of Pittsburgh
55	Arizona State University	52	University of Rochester
55	Brandeis University	56	Michigan State University
55	Georgia Institute of Technology	56	University of California, Riverside
55	Purdue University, West Lafayette	56	University of California, Santa Cruz
55	Stony Brook University, SUNY	56	Vanderbilt University
55	University of California, Santa Cruz	60	Northeastern University
55	University of Florida	60	University of Illinois, Chicago
55	University of Iowa	60	University of Notre Dame
55	University of Maryland, College Park	63	Iowa State University
55	University of Massachusetts, Amherst	63	University at Buffalo, SUNY
55	University of Oregon	63	University of Iowa
55	University of Southern California	63	University of Oregon
55	University of Utah	67	George Mason University
68	New York University	67	Oregon State University
68	Oregon Health and Science University	67	Syracuse University
68	Rutgers	70	Case Western Reserve University
68	Tufts University	70	College of William and Mary
68	University of California, Riverside	70	Colorado State University
68	University of Kansas	70	Naval Postgraduate School
68	University of Rochester	70	New York University
		70	Tufts University
		70	University of Delaware
		70	University of Maryland, Baltimore County
		70	University of Nebraska, Lincoln
		70	University of Tennessee, Knoxville
		70	University of Texas, Dallas
		70	Washington State University

Table S2. Career data set: Pooled cross-sectional model. The dependent variable is career achievement, measured as the natural logarithm of the Google Scholar citations, $\ln C_i$ as of 2017. The regression model is specified in Eq. (1) and estimated using standard OLS; there are 4,190 $\mathcal{F}_i$ (observations) for the pure CV model and 3,900 observations for the other two models that include network attributes, as in these cases we exclude from consideration disconnected $\mathcal{F}_i$ nodes. Natural logs were used to obtain variables that are approximately normally distributed. Thus, when the independent variable enters in ln, then β corresponds to the % change in C_i following a 1% change in the independent variable; in the case of the cross-disciplinarity fraction, β_χ represents the % change in C_i following a 0.01 shift increase in χ_i. The first column cluster shows the estimates using only standard CV variables. The combined *CV + Network* model demonstrates that $\mathcal{F}_i$ with larger χ_i correlate with higher net citation impact. For the combined model we also report the standardized beta coefficients – useful for comparing the relative strength of covariates within the regression. Standard errors were calculated using the clustered sandwich estimator, clustering on $\mathcal{F}_i$ age-cohort $y^0_{i,5}$ (based on 14 non-overlapping 5-year career birth year groups, e.g., 1940-1944, 1945-1950, etc.) to account for within-age-cohort correlation. Y indicates additional fixed effects included in the regression model.

	CV		**CV + Network**		**CV + Network [Standardized]**	
CV parameters						
Departmental rank, β_r	-0.052^{***}	(0.006)	-0.047^{***}	(0.005)	-0.056^{***}	(0.006)
Productivity (h-index), β_h	1.857^{***}	(0.020)	1.866^{***}	(0.022)	1.236^{***}	(0.015)
Total NSF funding, $\beta_{\$1}$	-0.005	(0.003)	-0.005	(0.003)	-0.036	(0.020)
# of NSF grants, β_{N1}	0.024	(0.013)	0.013	(0.014)	0.015	(0.015)
Total NIH funding, $\beta_{\$2}$	0.016^{***}	(0.003)	0.014^{***}	(0.002)	0.082^{***}	(0.014)
# of NIH grants, β_{N2}	-0.067^{***}	(0.015)	-0.061^{***}	(0.012)	-0.068^{***}	(0.014)
Network parameters						
PageRank Centrality, $\beta_{\mathscr{C}^{PR}}$			0.041	(0.019)	0.026	(0.012)
Cross-disciplinarity, β_χ			0.571^{***}	(0.073)	0.085^{***}	(0.011)
Discipline ($\mathcal{O}$) dummy	Y		Y		Y	
5-year cohort ($y^0_{i,5}$) dummy	Y		Y		Y	
Constant	1.492^{***}	(0.087)	1.668^{***}	(0.226)	7.609^{***}	(0.009)
n	$4,190$		$3,900$		$3,900$	
adj. R^2	0.883		0.882		0.882	

Standard errors in parentheses

* $p \leq .05$, ** $p \leq .01$, *** $p \leq .001$

Table S3. Career data set: Pooled cross-sectional model—robustness check. Parameter estimates for variants of the 'CV + Network' pooled cross-sectional models reported in table S2: (a) Model with PageRank centrality. (b) Model with betweenness centrality. (c) Model with degree centrality; (d) Model without the number of grants variables; (e) Model without the departmental rank variable. Results are not significantly different with respect to the primary covariate of interest, that is, cross-disciplinarity (β_χ).

	(a) $\mathscr{C}^{PR}$	(b) $\mathscr{C}^{B}$	(c) $\mathscr{C}^{D}$	(d) $\cancel{\beta_{N1}, \beta_{N2}}$	(e) $\cancel{\beta_r}$
CV parameters					
Departmental rank, β_r	-0.047^{***}	-0.042***	-0.044^{***}	-0.046***	
	(0.005)	(0.005)	(0.005)	(0.005)	
Productivity (h-index), β_h	1.866***	1.901***	1.848***	1.862***	1.892***
	(0.022)	(0.024)	(0.024)	(0.020)	(0.022)
Total NSF funding, $\beta_{\$1}$	-0.005	-0.004	-0.005	-0.003	-0.005
	(0.003)	(0.004)	(0.003)	(0.002)	(0.003)
# of NSF grants, β_{N1}	0.013	0.009	0.007		0.006
	(0.014)	(0.020)	(0.014)		(0.015)
Total NIH funding, $\beta_{\$2}$	0.014***	0.014***	0.014***	0.003*	0.013***
	(0.002)	(0.003)	(0.002)	(0.001)	(0.003)
# of NIH grants, β_{N2}	-0.061^{***}	-0.065***	-0.062^{***}		-0.059^{***}
	(0.012)	(0.013)	(0.012)		(0.014)
Network parameters					
PageRank centrality, $\beta_{\mathscr{C}^{PR}}$	0.041			0.042	0.057*
	(0.019)			(0.019)	(0.020)
Betweeness centrality, $\beta_{\mathscr{C}^{B}}$		-0.000			
		(0.006)			
Degree centrality, $\beta_{\mathscr{C}^{D}}$			0.052**		
			(0.016)		
Cross-disciplinarity, β_χ	0.571***	0.562***	0.530***	0.579***	0.555***
	(0.073)	(0.054)	(0.072)	(0.073)	(0.076)
Discipline ($\mathcal{O}$) dummy	Y	Y	Y	Y	Y
5-year cohort ($y_{i,5}^{0}$) dummy	Y	Y	Y	Y	Y
Constant	1.668***	1.192***	1.293***	1.671***	1.579***
	(0.226)	(0.083)	(0.069)	(0.226)	(0.225)
n	3900	3387	3900	3900	3900
adj. R^2	0.882	0.873	0.883	0.882	0.881

Standard errors in parentheses, listed below coefficient estimate.

* $p \leq .05$, ** $p \leq .01$, *** $p \leq .001$

Table S4. Career data set: Panel model on all faculty $\mathcal{F}$. Each column cluster reports the estimated coefficients for a specific model in which the dependent variable is the normalized citation impact of an individual article, $z_{i,p}$ belonging to faculty $\mathcal{F}_i$ - see Eq. (4). The first two column clusters correspond to a panel regression without $\mathcal{F}_i$ fixed effects, whereas the last two column clusters correspond to a panel regression with $\mathcal{F}_i$ fixed effects. Estimates in the second and fourth column clusters are calculated using standardized variables, where each 'beta' coefficient indicates the change in $z_{i,p}$ associated with a one standard deviation shift in the corresponding independent variable. The model without fixed effects incorporates time-independent author-level characteristics, i.e., adding to the specification of Eq. (2) the additional terms $[\beta_{\mathscr{C}^{PR}} \ln \mathscr{C}_i^{PR} + \beta_\lambda \ln \lambda_i + D(\mathcal{F}_i)]$. This is the reason why we only analyzed the 3,900 scholars connected within the network for which $\mathscr{C}_i^{PR}$ is defined; note that these additional variables are absorbed into β_i in the fixed effects model. The additional connectivity variable λ_i is the fraction of the total pollinators that are 'bridge' pollinators. Robust standard errors are shown in parenthesis, and X denotes time-independent variables absorbed by the fixed effects model. Y indicates additional fixed effects included in the regression model.

	No Fixed Effects	**No Fixed Effects [Standardized]**	**Fixed Effects**	**Fixed Effects [Standardized]**
Publication characteristics				
# of co-authors, β_a	0.284***	0.189***	0.312***	0.208***
	(0.00718)	(0.00479)	(0.00547)	(0.00365)
Career age, β_τ	-0.00547***	-0.0564***	-0.00949***	-0.0978***
	(0.000919)	(0.00947)	(0.00182)	(0.0187)
Cross-disciplinary indicator, β_I	0.126***	0.126***	0.145***	0.145***
	(0.0341)	(0.0341)	(0.0235)	(0.0235)
Network characteristics				
PageRank centrality, $\beta_{\mathscr{C}^{PR}}$	0.0440**	0.0284**	X	X
	(0.0142)	(0.00920)		
Bridge fraction, β_λ	0.334***	0.129***	X	X
	(0.0256)	(0.0099)		
Discipline ($\mathcal{F}$) dummy	0.00790	0.00790	X	X
	(0.0139)	(0.0139)		
Constant	0.451**	0.151	-0.293***	-0.0653**
	(0.142)	(0.102)	(0.0528)	(0.0202)
Year dummy	Y	Y	Y	Y
n	413,565	413,565	413,565	413,565
adj. R^2	0.055	0.055	0.036	0.036

Standard errors in parentheses below estimate.

* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$

Table S5. Career data set: Panel model on the $XD_{\mathcal{F}}$ faculty. Robustness check of panel model without and with fixed effects, implemented using only the 1,247 $\mathcal{F}_i$ with orientation $\mathcal{O}(\mathcal{F}_i) = XD_{\mathcal{F}}$.

	No Fixed Effects	No Fixed Effects [Standardized]	Fixed Effects	Fixed Effects [Standardized]
Publication characteristics				
# of co-authors, β_a	0.329***	0.220***	0.351***	0.234***
	(0.0123)	(0.00821)	(0.00880)	(0.00588)
Career age, β_τ	-0.00499**	-0.0514**	-0.00616*	-0.0635*
	(0.00181)	(0.0187)	(0.00253)	(0.0261)
Cross-disciplinary indicator, β_I	0.109***	0.109***	0.112***	0.112***
	(0.0328)	(0.0328)	(0.0234)	(0.0234)
Network characteristics				
Author centrality, $\beta_{\mathscr{C}}$	0.0526*	0.0340*	X	X
	(0.0265)	(0.0171)		
Bridge fraction, β_λ	0.319***	0.124***	X	X
	(0.0493)	(0.0191)		
Discipline ($\mathcal{F}$) dummy	-0.0383	-0.0383	X	X
	(0.0256)	(0.0256)		
Constant	0.217	-0.0773	-0.409***	-0.0685*
	(0.236)	(0.157)	(0.0778)	(0.0324)
Year dummy	Y	Y	Y	Y
n	166,621	166,621	166,621	166,621
adj. R^2	0.067	0.067	0.049	0.049

Standard errors in parentheses below estimate.

* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$

Table S6. Career data set: Panel model on the $XD_{\mathcal{F}}$ faculty with matched pairs. Robustness check of panel model without and with fixed effects, implemented using only the 53 $\mathcal{F}_i$ with orientation $\mathcal{O}(\mathcal{F}_i) = XD_{\mathcal{F}}$ who have at least 10 matched pairs of publications. Where possible, we matched each p with $I_{i,p}^{XD} = 1$ with a publication with $I_{i,p}^{XD} = 0$ from the same $\mathcal{F}_i$, having published within two years from each other, and featuring number of co-authors a_p that do not differ more than 20%.

	No Fixed Effects	No Fixed Effects [Standardized]	Fixed Effects	Fixed Effects [Standardized]
Publication characteristics				
# of co-authors, β_a	0.511***	0.375***	0.508***	0.373***
	(0.0451)	(0.0331)	(0.0453)	(0.0333)
Career age, β_τ	-0.00221	-0.0229	0.0474***	0.491***
	(0.00582)	(0.0603)	(0.00502)	(0.0519)
Cross-disciplinary indicator, β_I	0.133**	0.133**	0.135**	0.135**
	(0.0469)	(0.0469)	(0.0471)	(0.0471)
Network characteristics				
Author centrality, $\beta_{\mathscr{C}}$	0.217*	0.143*	X	X
	(0.0980)	(0.0648)		
Bridge fraction, β_λ	0.748**	0.222**	X	X
	(0.229)	(0.0682)		
Discipline ($\mathcal{F}$) dummy	-0.208	-0.208	X	X
	(0.116)	(0.116)		
Constant	-0.414	-1.538***	-2.268***	-0.307***
	(0.748)	(0.171)	(0.0980)	(0.0862)
Year dummy	Y	Y	Y	Y
n	1930	1930	1930	1930
adj. R^2	0.252	0.252	0.093	0.093

Standard errors in parentheses below estimate.

* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$