Science Advances
www.advances.sciencemag.org/cgi/content/full/1/1/e1400005/DC1

Supplementary Materials for

Systematic inequality and hierarchy in faculty hiring networks

Aaron Clauset,* Samuel Arbesman, Daniel B. Larremore

*Corresponding author. E-mail: aaron.clauset@colorado.edu

Published 12 February 2015, Sci. Adv. 1, e1400005 (2015)
DOI: 10.1126/sciadv.1400005

The PDF file includes:

Fig. S10. Prestige scores for the top 60 institutions for (A) computer science, (B) business, and (C) history.
Fig. S11. Centrality versus prestige rank for (A) computer science, (B) business, and (C) history departments, where centrality is defined as the mean geodesic distance (also known as closeness) divided by the maximum geodesic distance (diameter).
Fig. S12. Relative change in rank from doctoral to current institution for all Full, Associate, and Assistant Professors in (A) computer science, (B) business, and (C) history.
Fig. S13. Geographic structure of faculty hiring.
References (34–49)

Other Supplementary Material for this manuscript includes the following:
(available at www.advances.sciencemag.org/cgi/content/full/1/1/1400005/DC1)

Dataset 1: Business Faculty-Hiring Network Edges.
Dataset 2: Business Faculty-Hiring Network Vertex Attributes.
Dataset 3: Computer Science Faculty-Hiring Network Edges.
Dataset 4: Computer Science Faculty-Hiring Network Vertex Attributes.
Dataset 5: History Faculty-Hiring Network Edges.
Dataset 6: History Faculty-Hiring Network Vertex Attributes.

1 Supplementary Materials

1.1 Faculty hiring data collection

For each discipline (computer science, business, history), a list of PhD-granting institutions was identified. Institutions for business and history were taken from the *U.S. News & World Report* (US News) authoritative ranking most recent to the data collection (2012 and 2009 respectively). Institutions for computer science were taken from the authoritative Forsythe list, maintained by the Computing Research Association (`cra.org`). The computer science list thus included 23 Canadian institutions in addition to 182 American institutions, while the business and history lists are American-only. (A geographic analysis shows that Canadian computer science departments are sufficiently well-integrated into the U.S. system that including them does not bias our results; see SM Section 1.5.) Other international institutions were excluded from analysis in order to minimize the impact of language-based or national boundaries between hiring markets.

US News only provided numerical rankings for 112 business schools, with several hundred more listed as "unranked;" the ranked schools were selected to represent the business discipline. The complete US News list of 144 departments was selected for history. Computer science (CS) presented some ambiguities that required manual resolution to identify the target faculty at a given institution. For example, some universities have joint CS and electrical engineering (EE) departments and some institutions have both computer engineering (CE) and CS departments. Joint CS and EE departments were separated manually into CS and EE faculties, using institutionally given identifiers if possible or, as a last resort, by faculty research interests, with only CS faculty being included in the final listing. For institutions with multiple computing departments (e.g., CS and CE), the more CS-oriented department was chosen, except when the institution had a school or college of computing, e.g., Carnegie Mellon or Georgia Tech, in which case all member faculty were included. Faculty whose primary affiliation was with

	computer science	business	history
institutions	205	112	144
tenure-track faculty	5032	9336	4556
mean size	25	83	32
Full Professors	2400 (48%)	4294 (46%)	2097 (46%)
Associate Prof.	1772 (35%)	2521 (27%)	1611 (35%)
Assistant Prof.	860 (17%)	2521 (27%)	848 (19%)
female	15%	22%	36%
doctorates in-sample	87%	84%	89%
collection period	5/11 – 3/12	3/12 – 12/12	1/13 – 8/13

Table S1: Data summary for collected tenure-track faculty from each discipline. Doctorates in-sample gives the fraction of these tenure-track faculty whose doctorates were granted by institutions within our sampled set (out-of-sample faculty received their doctorates mainly from international institutions).

another department, i.e., courtesy appointments, were omitted. Non-tenure-track faculty, i.e., research faculty and instructors, were also omitted.

Faculty rostered within a particular discipline were considered to have been produced by that discipline. This choice was made in part because doctoral discipline was unavailable for a large number of faculty records, and even when discipline was available, we could not eliminate the plausible possibility that a rostered faculty whose doctorate was from outside the discipline was actually trained to some degree by within-discipline or affiliated faculty at the doctoral institution.

Genuine errors that may have been induced by this design were mitigated by two additional design choices. First, the disciplines chosen for study each have a fairly non-porous disciplinary boundary, with the large majority of rostered faculty being produced within the within-sample departments themselves. (Of these, business is the most porous and history is the least porous.) Second, the bootstrap uncertainty analysis described below provides a quantitative control on small biases created by the faculty-rostering assumption. So long as the number of such incorrect attributions is small—which the rest of the research design would tend to ensure—the overall results will be robust.

For approximately 10 faculty or less, the institution of the doctorate could not be determined, and these faculty were omitted completely. We note that a more complete study of all faculty in all departments would shed considerable light on the flow of faculty between disciplines, but we do not consider such effects here.

For each identified academic unit, a full listing of tenured or tenure-track faculty was obtained from the department's homepage (job titles equivalent to Assistant, Associate or Full Professor), and the following information was obtained for each: doctoral institution, current faculty rank, gender (male or female), and whether or not the faculty held one or more postdoctoral fellowships before becoming faculty. All information was collected manually from public data sources on the World Wide Web, typically from the faculty member's curriculum vitae or biography on their homepage. To reduce overall error rates, a uniformly random 20% of faculty in each unit were independently and manually verified (re-collected).

All records were then processed algorithmically for inconsistencies, coding errors, missing data, disambiguation, and other issues; all such errors were then corrected by hand, following the original protocol. Table S1 provides a summary of these data for each discipline, and Table S2 summarizes measures of inequality across disciplines.

We thank K. Parthan, S. Dhall, D. Mylsamy, P. Srivastava, S. Kamat, A. Yarlagadda, R. Daubert, A. Zizzi, K. Hargett, R.T. Guthikonda, R. Vardha, U. Lingala, and C. Sivakumar for help collecting these data, which took roughly 5000 person-hours of combined effort.

1.2 Minimum violation rankings

For each network, we extract a prestige hierarchy, defined as a ranking on the vertices that exhibits a minimum number of "violations," where a violation is an edge (u, v) in which the rank of v is better than the rank of u. Networks with strong prestige hierarchies will have only a small number of such violations.

	computer science	business	history
vertices	205	112	144
edges	4388	7856	4072
mean degree, $\langle k \rangle$	21.4	70.1	28.6
50% faculty from	18	16	8
Gini, $G\,(k_o)$	0.69	0.62	0.72
Gini, $G\,(k_i)$	0.32	0.25	0.26
$k_o/k_i > 1$	24%	36%	18%
ρ	0.88(2)	0.86(2)	0.91(2)
mean $\Delta\pi$	47.2(8)	27.1(3)	42.1(6)
π crossover	44 (36,61)	32 (30,49)	21 (15,34)
r^2_{USNews}	0.80	0.51	0.74

Table S2: Statistical measures of inequality by discipline: number of institutions required to cover 50% of disciplinary faculty; Gini coefficients for production (out-degree, k_o) and department size (in-degree, k_i); fraction of departments that are net faculty producers ($k_o/k_i > 1$); fraction ρ of placements that point down the prestige hierarchy π; average number of institutions down in ranking from doctoral to faculty institution; the prestige rank π at which institutions generally crossover from being net producers to net consumers of faculty, i.e., where $k_o/k_i \approx 1$; Pearson correlation coefficient between the prestige hierarchy π and US News ranking. Single parenthetical values give the uncertainty in the final digit; pairs give the 25–75% quantiles. Edge counts here are lower than Table S1 because faculty whose doctoral institution is outside our sample are excluded.

Let π denote a permutation of the n vertices in a directed network A, whose edges have positive weight, and let π_u denote the index of u within π. By convention, we call $\pi_u = 1$ the highest-ranked vertex. Finally, let $\pi(A)$ denote an adjacency matrix for A whose rows and columns have been reordered according to the permutation π.

A *minimum violation ranking* (MVR) is a permutation π that induces a minimum number of edges that point "up" the ranking (equivalently, a permutation that has the maximum number of edges that point "down," as in the main text). Such a permutation may be found by maximizing the net weight of unviolated edges, calculated as

$$S[\pi(A)] = \sum_{uv} A_{uv} \times \text{sign}[\pi(v) - \pi(u)] \ , \tag{1}$$

which subtracts the weight of all rank-violating edges, contained in the lower triangle (excluding

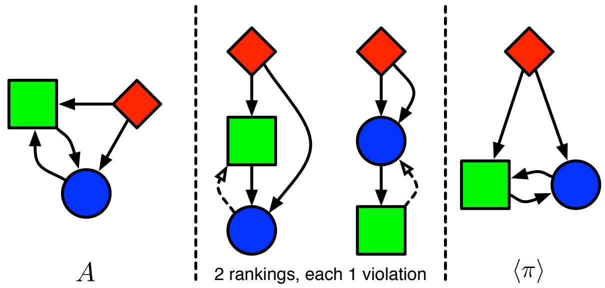

Figure S1: An example graph A, the two minimum violation rankings (MVRs) on these vertices, both with $S[\pi(A)] = 3$, and a "consensus" hierarchy, in which the position of each u is the average of all positions that u takes in the MVRs.

the diagonal) of the reordered adjacency matrix $\pi(A)$, from the weight of all non-violating edges.

Complex networks often exhibit multiple MVRs (*15*), in which many distinct vertex orderings produce the same smallest number of violations (Fig. S1). Instead of arbitrarily choosing one such ranking (*14*), we sample the MVRs using a stochastic optimization technique that is equivalent to a Markov chain Monte Carlo (MCMC) algorithm (*34*), with a Metropolis-Hastings acceptance function with zero temperature. At each step in our algorithm, a uniformly random pair of vertices is chosen and a ranking is proposed in which their ranks are swapped.[1] The choice of zero-temperature MCMC implies that all "neutral" or better proposals, i.e., those that do not increase the number of violations, are accepted, allowing the algorithm to sample the set of MVRs once it has converged on the minimum number of violations.

From the sampled ensemble of rankings for a given network, we compute the mean rank for each vertex, along with an uncertainty estimate for this score. To further incorporate uncertainty from the data themselves, we bootstrap the empirical data and compute the mean rank score across these replications (*35*). Under each bootstrap, we sample, with replacement, each

[1] In practice, higher-order exchanges, i.e., rotations of three or more vertices, appear not to produce fewer violations on these networks.

edge of the original faculty hiring network with equal probability. The resulting "consensus" hierarchy is a more robust estimate of the underlying ordering in the network. Furthermore, this resampling naturally simulates the underlying data generation process, and controls for modest-sized fluctuations in faculty hiring counts due to the particular timing of the data collection effort (Table S1).

The ensemble of MVRs also provides a basis for estimating uncertainty in subsequent analyses, e.g., the correlation between prestige rank and vertex centrality (*18, 36*), by computing the distribution of some quantity over members of the ensemble (for instance, Figs. S8–S11).

Applied to the the faculty hiring networks of computer science (CS), business (BS) and history (HS), the extracted consensus prestige hierarchies contain only a small number of violations: 12% in computer science, 14% in business, and 9% in history. A modest number of edges represent "self-loops" or institutions hiring their own graduates: 8% in computer science, 6% in business, and 5% in history. The presence or absence of these self-loops does not change the score of a ranking π under Eq. (1), nor do they substantially impact any subsequent analyses. Whenever a distinction between moving "up" or "down" a ranking is required, self-loops are considered to be the smallest possible move downward.

Figure S10 shows the resulting consensus rankings for the top 60 institutions in each discipline. Figure S3 shows the estimated variance (std. dev.) in score as a function of estimated mean score, illustrating a common convex shape in which uncertainty is lowest for the highest- and lowest-ranked institutions. This convex shapes reflects the fact that high-prestige institutions originate many edges, and that we may be fairly certain where they sit in the hierarchy, and similarly very low-prestige institutions originate very few edges, which tends to place them at the bottom of the ranking. Middle-ranked institutions are more ambiguous because the originate a modest number edges, and there are many orderings here that produce similar numbers of violations.

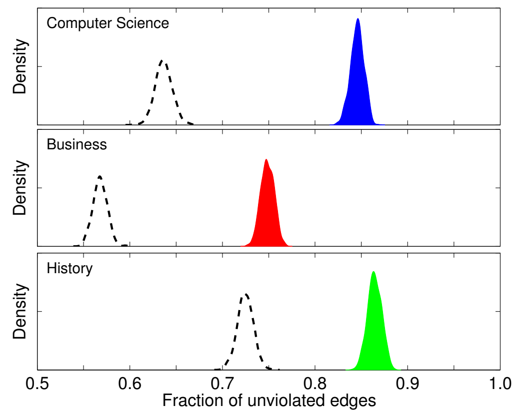

Figure S2: Bootstrap distributions (smoothed) for the fraction of unviolated edges ρ in the empirical data (filled) and in a null model (dashed), in which the in- and out-degree sequences are preserved but with the connections between them otherwise randomized, and those for the empirical data. The large gaps indicate that the extracted linear hierarchies are statistically significant.

Statistical significance of the extracted hierarchies

Random graphs with skewed degree distributions can exhibit hierarchies with a relatively small number of rank violations. We test the statistical significance of the extracted MVRs by comparing the number of unviolated edges in the empirical data against the same number for a random graph with the same degree sequence (*37*). In this random-graph model, the skewness in the degree distribution is preserved, but the empirical structure of the prestige hierarchy is removed.

We find that the estimated prestige hierarchies are highly statistically significant (each with $p < 10^{-5}$, Monte Carlo), and Figure S2 illustrates the strong separation between the hierarchical signal produced by the null model and that of the empirical data. In particular, the empirical hierarchies are 1.33, 1.32, and 1.19 times as strong (CS, BS, HS respectively) as the hierarchies for a random graph with the same degree distribution.

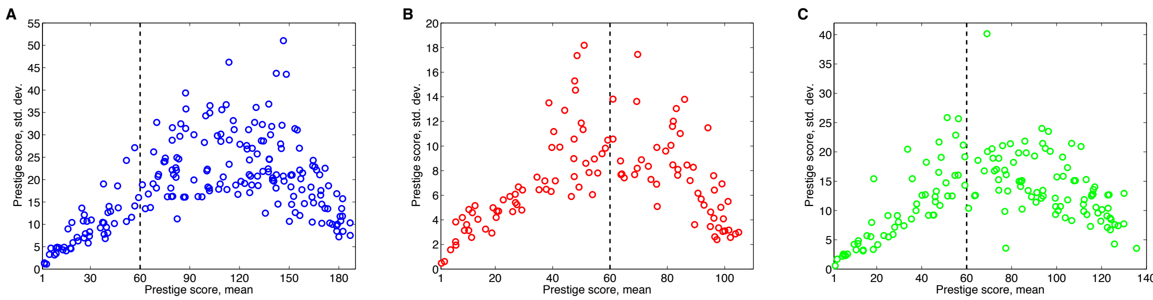

Figure S3: Prestige uncertainty versus prestige, shown as the standard deviation of the estimated distribution versus the distribution mean, for (**A**) computer science, (**B**) business, and (**C**) history. The vertical line demarcates the top 60 institutions (Fig. S10). The common convex shape illustrates relatively low uncertainty for very high or very low prestige institutions, with larger uncertainty for intermediate values of prestige, particularly below the top-60 line.

Rank-changes as a function of doctoral prestige

Under each hierarchy, the change-in-rank for each edge (u, v) is $\Delta\pi = \pi_u - \pi_v$. Figure S4 shows scatter plots of these rank changes by discipline, as a function of doctoral rank π_u. In accordance with the estimated fraction of edges pointing down each hierarchy, the vast majority of plotted points are below the $\Delta\pi = 0$ line. (Note: because 1 is the highest rank, a movement down the hierarchy implies $\Delta\pi > 0$.)

Although the distribution of rank changes varies continuously with prestige, a useful illustration is to compare the changes-in-rank for faculty originating in the top 15% versus those in the remaining institutions (Fig. 3, main text). For these faculty, the median rank-change is 3–18% lower than the median for faculty originating from less prestigious institutions. (In contrast, their mean rank-change is 13–25% greater, largely because a moderate portion of these faculty moved to the very lowest-ranked institutions.) This effect is smallest in history, which also has the largest inequality in faculty production (Fig. 2, main text).

Figure S5 shows the rank-change distributions for male and female faculty, for the same division of institutions (top 15% versus the rest), as discussed in the main text. The difference

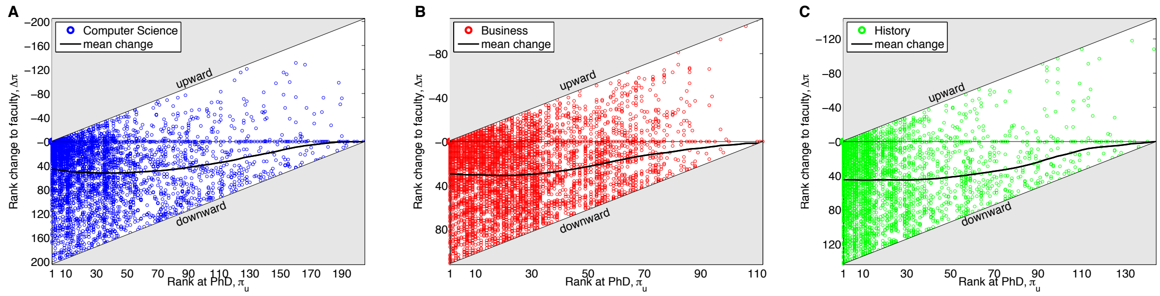

Figure S4: Changes in rank from doctoral institution u to faculty institution v, for each edge (u, v) in (**A**) computer science, (**B**) business, and (**C**) history. Solid black line shows the (smoothed) mean rank-change as a function of doctoral rank. Grey regions indicate forbidden changes in rank, which would take a faculty above rank 1 (highest) or below rank n (lowest).

between male and female distributions is statistically significant when data are combined across disciplines (KS test on normalized rank changes $\Delta\pi/n$, $p < 10^{-3}$). Differences are not statistically significant when male and female distributions are considered within a discipline. The difference likely reflects the relatively small number of women in each discipline.

The median change for men is 21–35 ranks downward, while for women it is 23–38. Differences by gender are greatest for graduates of the most prestigious institutions in computer science and business, where median placement for women graduating from the top 15% of units is 12–18% worse than for men. In contrast, we find no gender difference in median placement for history, which may reflect history's stronger hierarchy. (Similar results hold for other definitions of the "top" group; see Fig. S6)

Figure S7 shows the rank-change distributions for faculty who held at least one postdoctoral research position versus those that held none, for the same division of institutions. The overall difference between these distributions is statistically significant in business (KS test, $p < 10^{-6}$), but not in computer science or history ($p = 0.41$ and $p = 0.12$ respectively). The latter null results may be due in part to the small number of individuals who have held postdoctoral positions (16% in computer science, 2% in business, and 3% in history). Examining the medians,

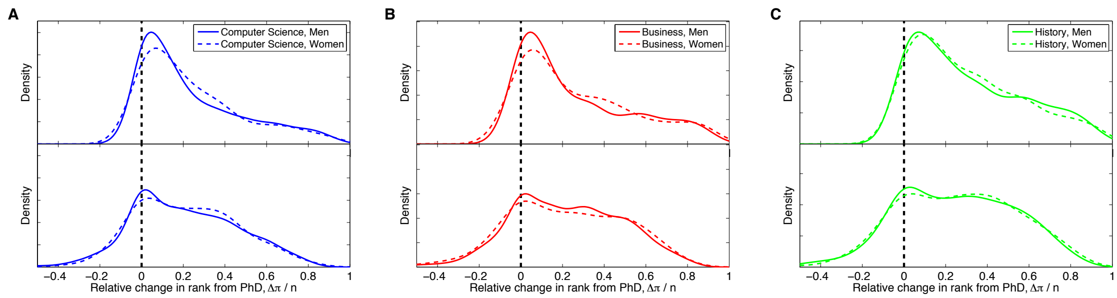

Figure S5: Changes in rank from doctoral institution u to faculty institution v, for each edge (u, v) in (**A**) computer science, (**B**) business, and (**C**) history, divided by male versus female faculty for u in the top 15% of institutions (top panels) or in the remaining institutions (bottom panels). These distributions are very similar, but still slightly different, with women from the top 15% of institutions placing slightly worse than men from the same institutions.

we observe a small improvement in placement for individuals who have held postdocs, and note that in every case, the proportion of individuals moving upward in the hierarchy is slightly larger for the postdoc case.

Finally, it is interesting to note that because higher-prestige correlates with greater net faculty production ($r^2 = 0.38$–0.59, $p < 10^{-10}$; Fig. S8D), an average step down in prestige corresponds to a reduction in unit-level faculty production by a factor of 2.0–3.8. Thus, at the individual level, a typical professor can expect to supervise 2–4 times fewer new within-discipline faculty than did their own doctoral advisor.

Mechanisms for prestige hierarchies

One plausible mechanism for the origin of a prestige hierarchy is a simple positive feedback loop between prestige and successful faculty placement. In this way, high-prestige institutions would more easily maintain their relative position in the hierarchy by virtue of placing a large number of faculty at lower-ranked institutions. That being said, small-producing institutions can rank highly if their graduates place at highly-ranked institutions. In computer science, Caltech is particularly notable for this, having a relatively small number of graduates placed as faculty,

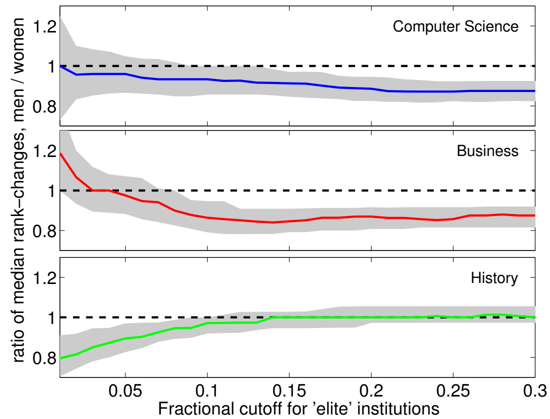

Figure S6: Ratio of the median change-in-rank, from doctoral institution u to faculty institution v, for men versus women, for faculty receiving their doctorate from the "most prestige" institutions, showing that elite women tend to place below their male counterparts in computer science and business (ratio < 1). Grey envelope shows the 10–90% bootstrap quantiles.

many of whom are placed at other prestigious institutions.

In contrast, lower-ranked institutions would have difficulty improving their standing because their smaller number of faculty candidates compete for a small number of positions with a very large number of graduates produced by more prestigious institutions. In fact, under this definition of institutional prestige, the primary way to improve a ranking is to successfully place faculty at higher-ranked institutions. Hiring faculty from more highly-ranked institutions only directly contributes to the prestige of the other institutions, not the institution doing the hiring.

As a result, faculty who successfully place above their doctoral rank are thus of particular interest, and would provide an instrumental approach to testing this hypothesis over time.

1.3 Alternative methods for ranking institutions

Here we describe and define alternative methods for ranking institutions, which work either by using information embedded in the faculty hiring network or by using features of the institutions, as in the US News and NRC rankings. Across methods, ties are broken uniformly at

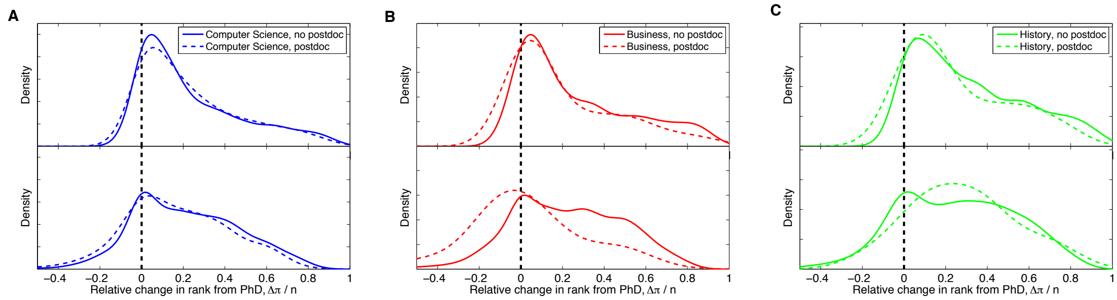

Figure S7: Changes in rank from doctoral institution u to faculty institution v, for each edge (u, v) in (**A**) computer science, (**B**) business, and (**C**) history, divided by faculty who have held one or more postdoctoral positions versus those that held none, for u in the top 15% of institutions (top panels) or in the remaining institutions (bottom panels). The most notable differences between these distributions occurs in the slightly greater upward mobility of individuals with postdocs.

random. Across disciplines, we find that network-based measures like degree centrality, eigenvector centrality, geometric centralities and net faculty production correlate strongly with the minimum violation ranking (Fig. S8, and statistics shown in Table S2), in agreement with a strong core-periphery structure (*17*). This finding is further reinforced by examining other aspects of the network, e.g., the minimum-spanning trees rooted at each vertex, and the smooth increase in mean distances along these shortest-path trees as a function of the prestige of u (Fig. S11).

Random

As a baseline ranking scheme, in which rank is uncorrelated with either network or institutional features, we take a uniformly random permutation of the n vertices.

Degree centrality

The simplest measure of importance in a network is the number of connections in which some vertex participates, i.e., the degree. We consider three such measures, the out-degree (faculty production k_o) the in-degree (faculty consumption k_i), and the net degree (ratio of production to consumption k_o/k_i). Out-degree has been used previously in a single-discipline study of faculty hiring in law (*27*). From the adjacency matrix A, in which $A_{uv} = 1$ indicates the directed relationship of a graduate of u being faculty at v (and is otherwise 0), these are defined mathematically as

$$k_o = \sum_u A_{uv} \qquad\qquad k_i = \sum_v A_{uv} \quad . \tag{2}$$

Eigenvector centrality

Eigenvector centrality is a recursive measure of vertex "importance" in a network. Loosely speaking, a vertex with high eigenvector centrality is one to which other vertices with high eigenvector centrality connect (*18, 19, 36*). Variations of this measure have been used previously in single-discipline studies of faculty hiring to rank institutions in sociology, political science, organizations science, economics, law and mathematics (*8, 25–31*). These studies did not compare the induced rankings against alternatives, except for mathematics, which showed that eigenvector centrality scores correlate with authoritative rankings by US News and NRC (*25*).

Mathematically, eigenvector centrality scores are computed from the eigenvectors of the adjacency matrix A

$$\gamma\,\mathbf{x} = (A + \epsilon)\,\mathbf{x} \ \ , \tag{3}$$

where $\mathbf{x}$ is the vector of eigenvector centralities, γ is the largest eigenvalue, and ϵ adds a small amount of noise to each entry in A. The eigenvector centrality, denoted λ in the text, of vertex u is the uth entry of $\mathbf{x}$.

The Perron-Frobenius theorem provides that for this calculation to have a solution, the adjacency matrix A must represent a single connected component, in which each vertex is reachable from any other vertex by some path (*19*). The faculty hiring matrix, being a directed network, does not necessarily fulfill this requirement, either naturally or under a bootstrap. Adding a small amount of noise of size $\epsilon = O(1/n)$ to each entry in A guarantees a valid solution.

Geometric centrality

Alternatively, important vertices in the network may be identified via geometric means (*19, 36*). From this perspective, important vertices are positioned more closely to all other vertices, with distances measured by geodesic (shortest) paths. We use two related versions of geometric centrality: harmonic centrality h_c and mean-geodesic distance c (also called "closeness" centrality). Mean geodesic distance has been used previously in a single-discipline study of faculty hiring in law (*27*).

Let d_{uv} be the length of the shortest path from vertex u to vertex v. If no path exists, then $d_{uv} = \infty$. The harmonic centrality (*19, 38*) for vertex u is

$$h_c = \frac{1}{n-1} \sum_{u \neq v} \frac{1}{d_{uv}} \quad , \tag{4}$$

while the mean geodesic distance c for vertex u is

$$c = \frac{1}{n} \sum_{v} d_{uv} \quad . \tag{5}$$

In both cases, edge weights in the adjacency matrix are ignored. When some pair u, v has no path between them, closeness c is undefined for all vertices. In our subsequent analysis, when c is undefined, a random ranking (see above) is substituted.

Authoritative rankings

The most commonly referenced institutional rankings are those by the US News (*39*) and the NRC (*9, 40*). These methods use a common "weighted sum" approach: (i) a set of numerical features are established ("proxy" variables for behaviors of interest, e.g., selectivity, "reputation" and financial resources), (ii) each feature is assigned a weight, meant to represent its importance in determining institutional quality, and (iii) institutions are ordered according to their weighted sum on these features.

This approach is widely criticized for poor methodological structure (*10, 41–46*) and accuracy (*11, 47*). Critics argue that (i) the chosen features and weights improperly emphasize variables associated with educational "inputs" (e.g., wealth [funding, alumni giving, etc.]), fame (reputation among peer institutions), and exclusivity (how many applicants are rejected), (ii) variables associated with educational outcomes like knowledge mastery and job placement are omitted, and (iii) the rankings are often based on incorrect or outdated data. Furthermore, the input data are often self-reported by the ranked institutions, and are thus open to some degree of manipulation.[2] Despite such issues, these rankings are a standard evaluative measure of institutions of higher education and thus serve as a critical reference point for comparison against network-based methods.

In each discipline, the US News ranking most recent to the data collection was used (Computer Science 2010, Business 2012, and History 2009). The National Research Council does not evaluate business schools as a whole, and thus no NRC ranking is available for this discipline. Similarly, the computer science community rejected as fundamentally inaccurate the 2010 NRC ranking for computer science (*47*); thus, we used the most recent NRC ranking of CS prior to 2010, which was published in 1995. Finally, we used the 2010 NRC S-ranking for history; when multiple departments at a single institution appeared, we selected the one closest to a general history department.[3] Institutions within our sample that were not assigned a rank in a particular authoritative ranking, along with institutions that were included but which were

[2]In contrast, rankings based on faculty hiring networks are derived from publicly available data, e.g., curricula vitae or résumés, which are less easily manipulated by prestige-seeking administrations (*10, 43, 44*). For an institution to improve its standing in a MVR, it would be incentivized toward effective job placement. Hiring networks alone, however, provide an incomplete set of incentives because they omit information on scholarly activity, e.g., citations (*32*), subsequent to or independent of placement. We note that such information could also be evaluated using network-based tools.

[3]Specifically, Brandeis University's History department was assigned a rank of 7, the rank of its American History PhD program, rather than 61, the rank of its Comparative History program.

Computer Science	π	US News	NRC	λ
π	—	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)
λ	0.80 (205)	0.74 (153)	0.81 (103)	—
Business	π	US News	NRC	λ
π	—	0.51 (111)	—	0.85 (112)
US News	0.51 (111)	—	—	0.49 (111)
NRC	—	—	—	—
λ	0.85 (112)	0.49 (111)	—	—
History	π	US News	NRC	λ
π	—	0.74 (143)	0.32 (122)	0.72 (143)
US News	0.74 (143)	—	0.39 (123)	0.65 (143)
NRC	0.32 (122)	0.39 (123)	—	0.32 (122)
λ	0.72 (143)	0.65 (143)	0.32 (122)	—

Table S3: Pearson correlation coefficients r^2 between pairs of rankings. Parenthetical values indicate the number of institutions present in both rankings. All correlation coefficients are statistically significant ($p < 10^{-9}$). The NRC does not rank business schools.

given an "unranked" status, were added and assigned the lowest rank allowed (breaking ties randomly).

Table S3 gives the Pearson correlation coefficients for pairwise comparisons of the authoritative rankings with each other and with both the prestige hierarchy and the eigenvector centrality alternative measure. Notably, the prestige rankings correlate roughly as well with the authoritative rankings as those do with each other. This indicates that the prestige hierarchy captures at least as much information about these institutions as these authoritative rankings.

1.4 Quantifying placement accuracy

Given a ranking on the vertices, we quantify its accuracy for faculty placement by converting it into a non-parametric probabilistic model of edge existence and evaluating its ability to assign high likelihoods to observed placements, when given only the doctoral institution's rank. By estimating such a model on the complete faculty hiring data and then measuring its accuracy on

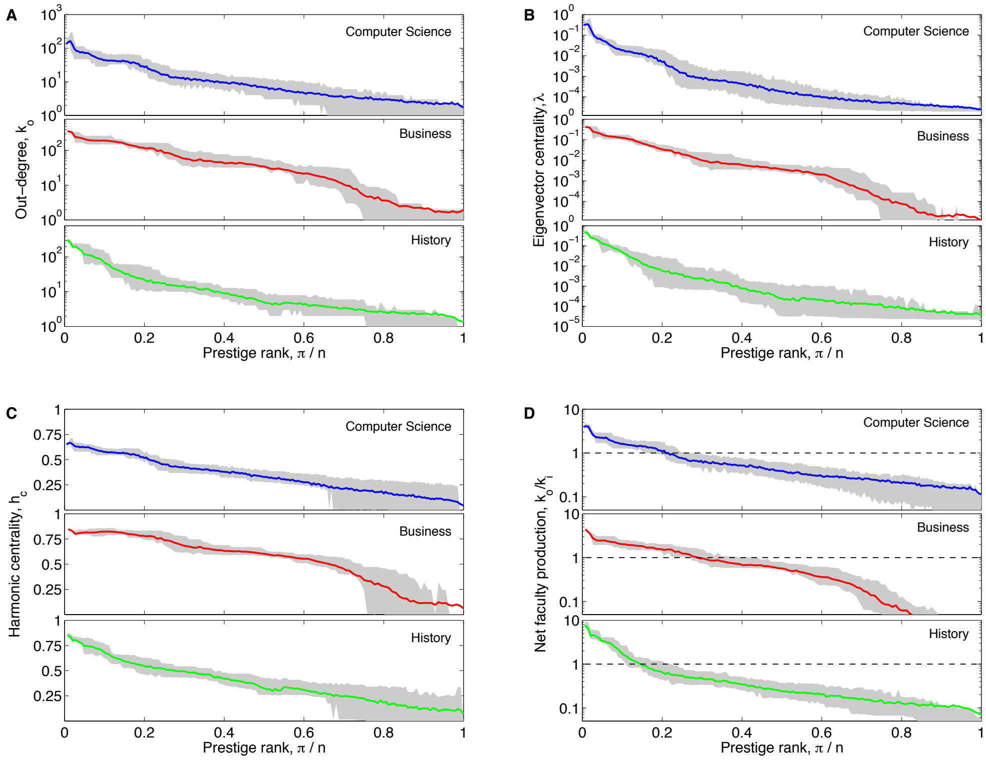

Figure S8: Centrality measures versus prestige rank. (**A**) Degree centrality (out-degree k_o), (**B**) eigenvector centrality λ, (**C**) harmonic centrality h_c, and (**D**) net degree (k_o/k_i). (Closeness centrality is shown in Fig. S11.) Solid lines show the mean relationship under a bootstrap of the faculty placement data, and the grey envelope shows the 25–75% bootstrap quantiles; dashed line in panel **d** shows the line separating net producers from net consumers of faculty. The decreasing function in each case indicates that each quantity correlates strongly with prestige.

a subset of the same data set, our placement accuracies are a measure of model goodness-of-fit. We characterize the model's accuracy on assistant professors alone because these are the most recent placements and thus likely most indicative of the current faculty hiring market.

Let π be a ranking on the vertices (derived according to any of the measures in SM Section 1.3), and let $\pi_u \in [1, n]$ be the rank of vertex u. Each Assistant, Associate or Full Professor corresponds to a particular directed edge (u, v) in the network, where u is the doctoral institution and v is the current institution for this professor. Thus, each edge represents a change in

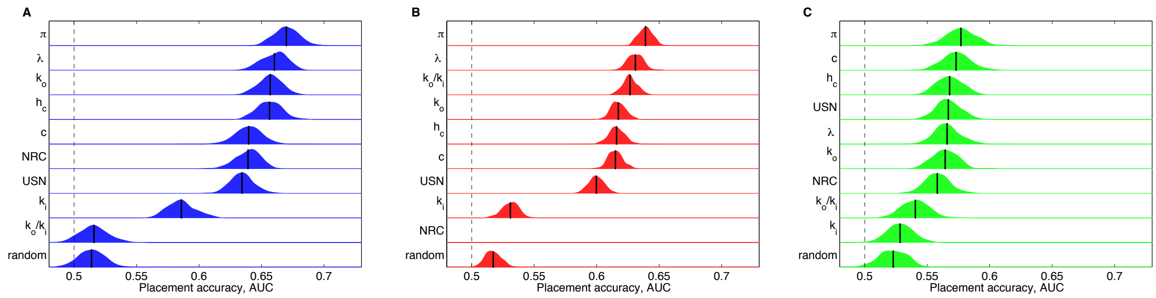

Figure S9: Placement accuracy for assistant professors. Results for (**A**) computer science, (**B**) business, and (**C**) history, with seven network rankings, prestige π, eigenvector centrality λ, out-degree k_o, harmonic centrality h_c, closeness centrality c, in-degree k_i, net-degree k_o/k_i, and the most recent authoritative rankings by *U.S. News & World Report* and the National Research Council. Across disciplines, prestige ranking (see text) outperform all other rankings. (Note: the NRC does not rank Business schools.)

rank $\Delta\pi = \pi_v - \pi_u$ from doctorate to faculty position. (Equivalently, we let a single edge (u, v) be assigned a weight equal to the number of faculty moving from u to v.)

From each of the empirically observed changes in rank, we make a non-parametric estimate of $\Pr(\Delta\pi/n)$, the distribution of relative rank changes induced by π on the placement network. Normalization by n ensures that this distribution is defined only on the interval $[-1, 1]$, and thus represents a model of rank changes regardless of the rank of the originating vertex u. To the empirically observed relative changes in rank, we apply a non-parametric Gaussian kernel (*48*) to produce a smoothed histogram, in increments of $1/n$ on the range $[-1, 1]$. This density function quantifies the particular ranking's estimate of the relative likelihood of an edge connecting a vertex of rank π_u with some other vertex of rank π_v. Figure S12 shows the resulting relative rank-change distributions for the ranking schemes used in this study (see SM Section 1.3), for each discipline.

Given a distribution, the placement accuracy of the particular ranking measure is directly measured by first taking the point-of-origin u for each of the directed edges for assistant professors and then ranking the $n - 1$ candidates for v by their estimated likelihood according to

$\Delta\pi/n = (\pi_v - \pi_u)/n$. If an institution produced multiple assistant professors, each is given the same ordering of predicted placements as there is no other information available to distinguish them. (This by itself serves to limit the maximum AUC score, as it is uncommon for multiple faculty originating at some u to both place as faculty at the same v.) From these predictions, a receiver operating characteristic (ROC) curve may be constructed, which quantifies how the true positive rate varies as a function of the false positive rate. We record the area under this curve (AUC) as a single measure of the overall accuracy, and takes a value between 0.5 (no better than chance) and 1 (perfect accuracy).

We estimate the variance in the AUC for placement accuracy via a nonparametric bootstrap of the input data. Specifically, we first sample with replacement from the set of faculty to construct an adjacency matrix with similar structural statistics. For each bootstrap, we recompute each of the possible rankings on the placement matrix and then measure their accuracy using the above procedure. Across disciplines, the MVR prestige hierarchy is always the most accurate (Fig. S9). Although some alternative measures are only slightly less accurate, the relative ordering of these alternatives varies by discipline, indicating that none is a good general predictor of placement, while prestige is.

1.5 Prestige and geography

For each discipline, a regional label can be assigned to each institution, according to a major U.S. Census region (Northeast, Midwest, South, or West) or Canada. Using these labels, we consider the geographic distribution of institutions with respect to their prestige.

Regional representation among high-prestige institutions

Network visualizations (force-directed layout) of the faculty hiring networks for each discipline, for all tenure-track faculty together and then again by faculty rank (Full, Associate, or

	Northeast	Midwest	South	West	Canada	All
Northeast	**0.119**	0.053	0.074	0.055	0.022	0.322
Midwest	0.031	**0.067**	0.061	0.026	0.011	0.196
South	0.025	0.027	**0.083**	0.024	0.006	0.166
West	0.049	0.033	0.043	**0.073**	0.011	0.209
Canada	0.006	0.005	0.005	0.005	**0.085**	0.107
All	0.229	0.185	0.267	0.184	0.135	

Table S4: Computer science, fraction of faculty who received their doctorate in one region (row) and are employed as faculty in another (column). Within-region movements are highlighted in bold.

Assistant), show that high out-degree departments, which also tend to be the high-prestige departments, are located in the centers of these network and are unequally distributed across these regions (Fig. S13A).

Although the variation appears to be continuous, again to illustrate the difference between high-prestige and other institutions, we divide institutions into "core" and "periphery," which we arbitrarily define as the top 15% of institutions by prestige and the remainder, respectively. We then test the extent to which regions are over- or under-represented among the high-prestige group in each discipline.

Across disciplines, we find that institutions in the Northeast are highly over-represented in the high-prestige group (relative to a null model in which a uniformly random set of institutions is chosen to be the core), while institutions from the South are highly under-represented (Fig. S13B). On average, the high-prestige group includes 23% more institutions than expected in the Northeast, and 22% fewer than expected from the South. This pattern is strongest in business, with the South including 35% fewer than expected institutions in the high-prestige group. These patterns are reversed for the remainder of institutions (lower prestige), which includes fewer than expected institutions from the Northeast and more than expected institutions from the South. The West was generally slightly better represented than expected, although not in a statistically significant way. In computer science, Canada is slightly under-represented.

	Northeast	Midwest	South	West	All
Northeast	**0.162**	0.057	0.069	0.041	0.328
Midwest	0.072	**0.090**	0.089	0.045	0.296
South	0.039	0.034	**0.116**	0.024	0.212
West	0.045	0.023	0.042	**0.053**	0.163
All	0.317	0.205	0.316	0.163	

Table S5: Business, fraction of faculty who received their doctorate in one region (row) and are employed as faculty in another (column). Within-region movements are highlighted in bold.

	Northeast	Midwest	South	West	All
Northeast	**0.146**	0.096	0.111	0.057	0.411
Midwest	0.055	**0.085**	0.075	0.035	0.250
South	0.021	0.030	**0.082**	0.011	0.144
West	0.041	0.042	0.055	**0.059**	0.196
All	0.262	0.253	0.324	0.161	

Table S6: History, fraction of faculty who received their doctorate in one region (row) and are employed as faculty in another (column). Within-region movements are highlighted in bold.

This finding indicates that institutional prestige is not independent of geography. Such a pattern likely reflects two well-known historical facts: (i) the oldest, and most entrenched institutions, e.g., the Ivy League, are located in the Northeast, and (ii) the South was largely an agriculture-based economy until after World War II and thus had substantially fewer resources to invest in higher educational institutions and invested much later in time than the Northeast. The appearance of these historical patterns in current data suggests a strong role for positive feedback mechanisms.

Regional faculty hiring patterns

To quantify the flow of faculty among these geographic regions, we tabulated the fraction of faculty who received their doctorate in region s and are now faculty in region t. Tables S4–S6 give the resulting flows, along with the fractional flow out of each region (given in the All column) and the fractional flow into each region (given in the All row).

Across disciplines, we find a common pattern that a plurality of tenure-track faculty received

Q_r	Computer Science	Business	History
Northeast	0.070	0.102	0.069
Midwest	0.042	0.041	0.038
South	0.046	0.057	0.039
West	0.035	0.027	0.027
Canada	0.071	–	–
Q	0.264	0.226	0.173

Table S7: Contributions Q_r by geographic region to total network modularity score Q, within each discipline.

their PhDs in the Northeast, reflecting this region's over-representation within the high-prestige group (Fig. S13B). Similarly, across disciplines, a plurality of tenure-track faculty are working in the South (indicated by the largest element in the "All" row of each table), with the exception of business, in which a nearly equal fraction work in the Northeast and the South. History, however, presents a particularly extreme situation: in this discipline, for both the South and the Midwest, a greater share of faculty working in these regions received their doctorates from the Northeast than from within the region. In no other discipline, and for no other regions, are within-region faculty a minority, which reflects the extreme concentration of history faculty production among Northeastern elite institutions.

In each of these matrices, the diagonal represents the fraction of these faculty who are both trained and employed in the same region, and in every case, these fractions are greater than the flows to other regions, generally by a factor of 2–6. The presence of such within-region flows can create community or modular patterns in the network (*49*). We quantify the extent to which regions form distinct hiring markets via the *modularity* score Q for each matrix. This measure quantifies the extent to which these flows tend to circulate within each region after controlling for the unequal production rates (*49*). The modularity is defined as

$$Q = \sum_r e_r - a_r^2 \ , \tag{6}$$

where e_r is the fraction of edges in the network that originate and terminate in region r and a_r

is the fraction of edges with at least one endpoint in region r. A value of $Q \approx 0$ indicates no more modular structure than we would expect at random, while a value close to unity indicates the presence of strong modular structure.

Here we find modularity scores in the range 0.17–0.26 (Table S7), which are not large by most standards in network science, but do indicate the presence of some modularity or regional substructure in the faculty hiring network. Across these disciplines, history is the least modular, while computer science is the most.

Each network-level score can be further decomposed into the relative contribution by each region $Q_r = e_r - a_r^2$, which gives an indication of the relative insularity of each region. Overall, this analysis shows that the Northeast is generally the most insular of the regions. This insularity is tied by Canada in computer science, indicating that Canada's CS faculty market is no more distinct from the North American market as a whole than is the Northeast's market. This fact supports the inclusion of the 23 Canadian CS departments in the overall analysis.

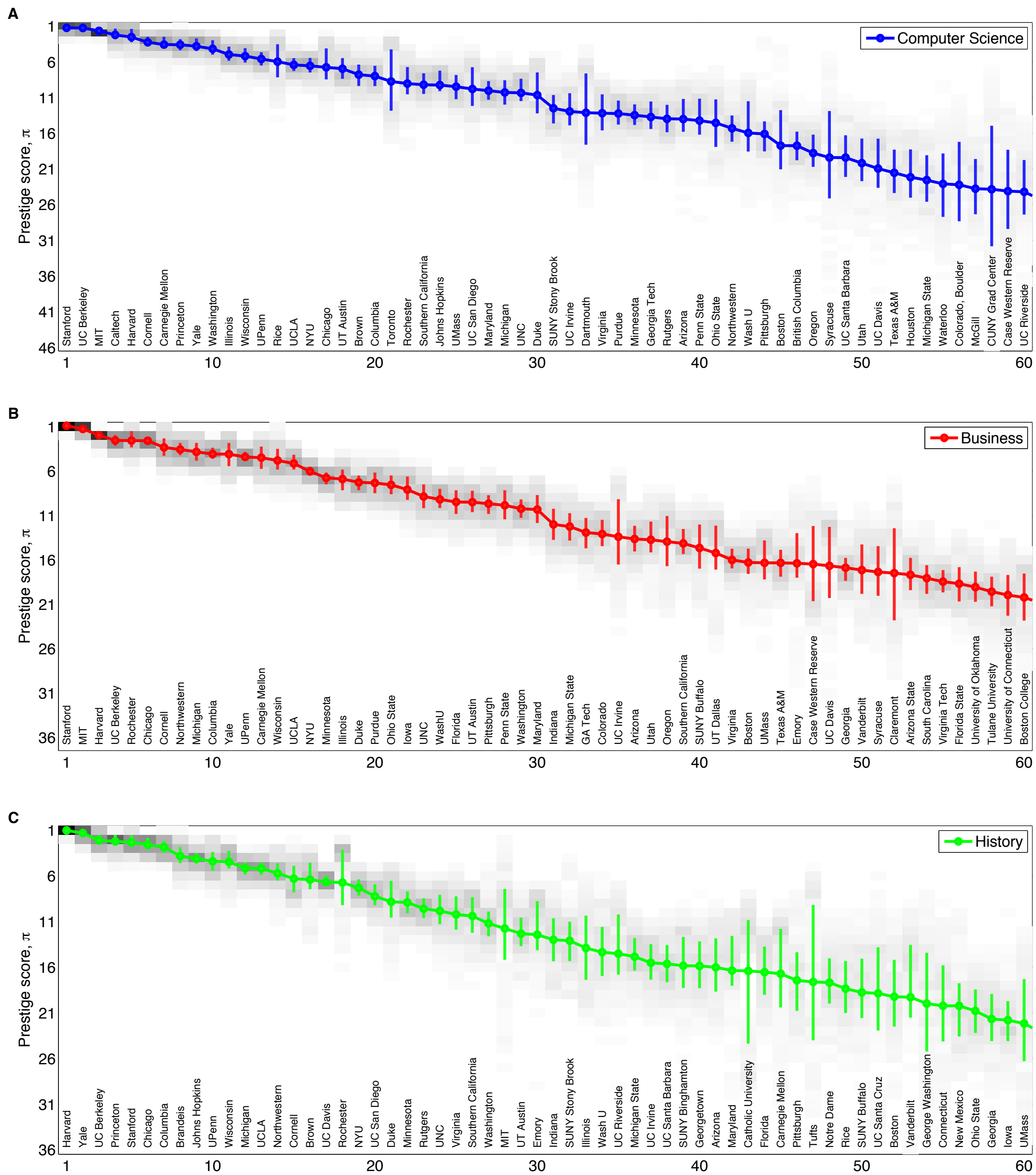

Figure S10: Prestige scores for the top 60 institutions for (**A**) computer science, (**B**) business, and (**C**) history. For each institution, its distribution of prestige scores $\Pr(\pi)$ is shown in greyscale (darker squares are greater density), on top of which is overlaid a vertical line showing the 25–75% quantile. A line connects the means of each distribution.

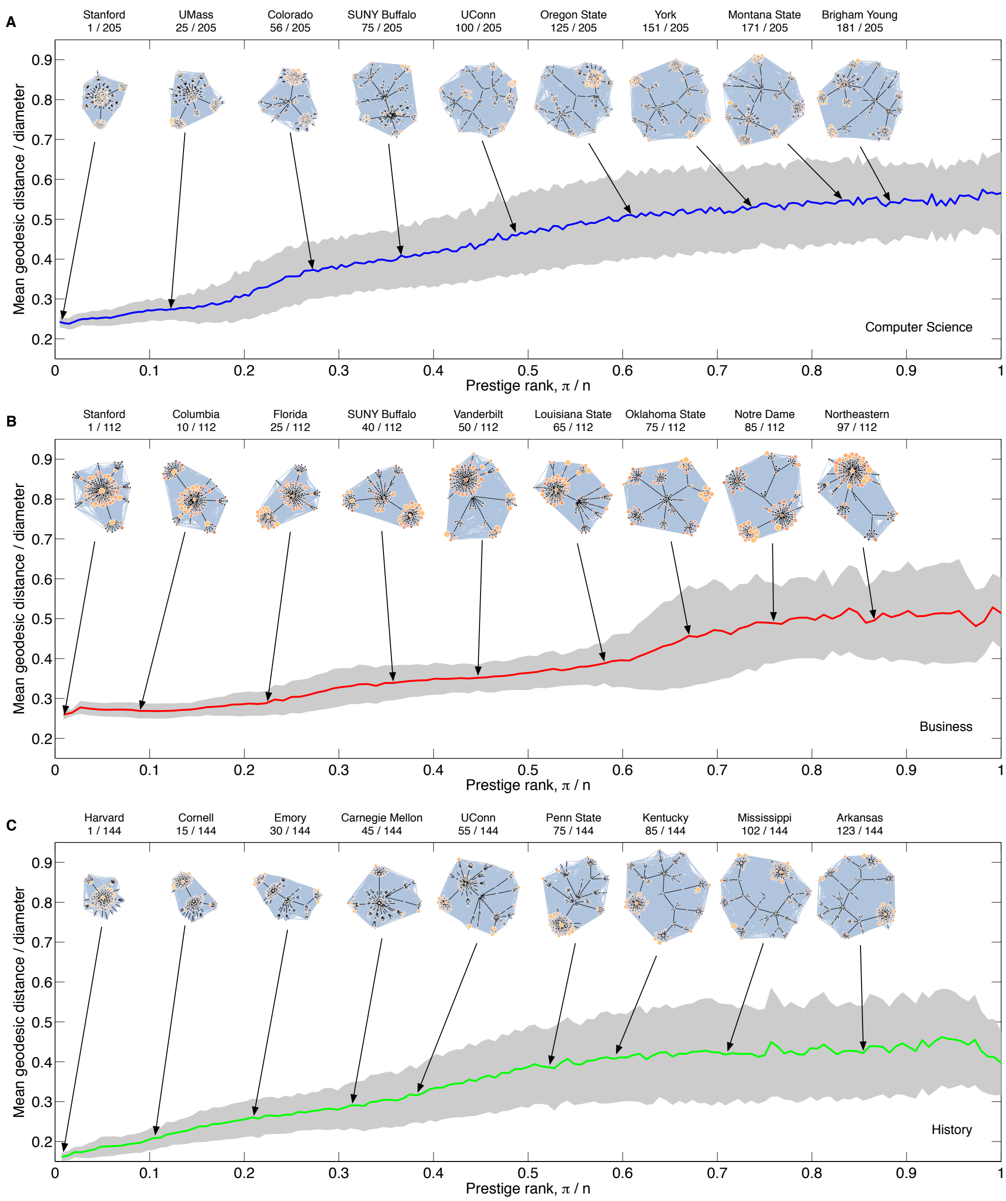

Figure S11: Centrality versus prestige rank for (**A**) computer science, (**B**) business and (**C**) history departments, where centrality is defined as the mean geodesic distance (a.k.a. closeness) divided by the maximum geodesic distance (diameter). Shaded region indicates the 25–75% quantiles around the mean (line). Network visualizations show the minimum spanning trees on the faculty hiring network, centered at selected vertices, illustrating the tendency for higher-prestige departments to be more centrally located within the network. (Vertex color is lighter for more prestigious institutions, and vertex size is proportional to out-degree.)

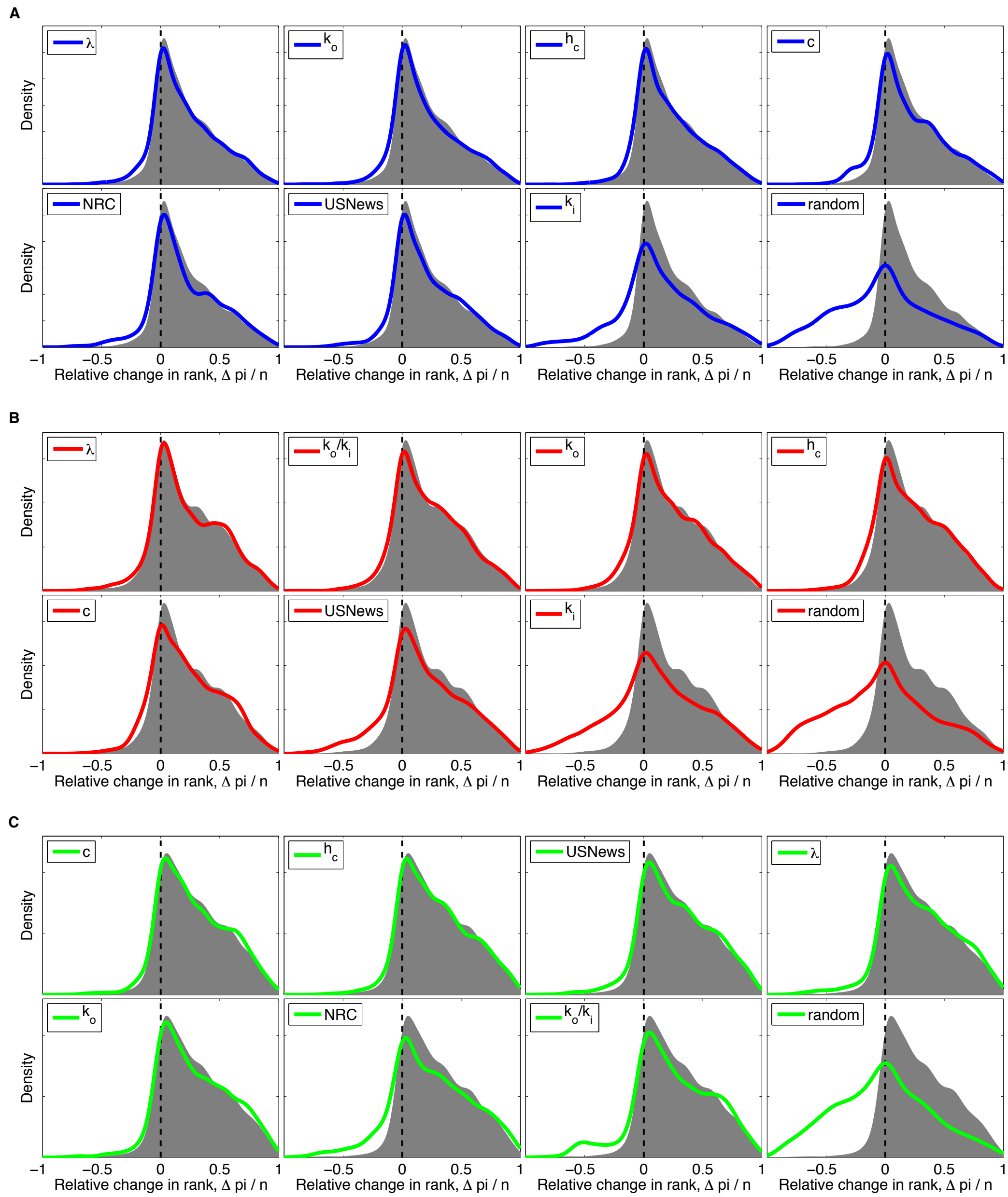

Figure S12: Relative change in rank from doctoral to current institution for all Full, Associate and Assistant Professors in (**A**) computer science, (**B**) business, and (**C**) history. Each panel shows the rank-change distribution for a different ranking method (see text), with the distribution induced by the minimum violation ranking (MVR) method shown in shaded grey (to provide a contrast with the most accurate model), and panels are given in decreasing order of the method's placement accuracy (see Fig. S9). In all cases, the MVR method produces a more concentrated and more right-skewed distribution, which produces more accurate placement predictions.

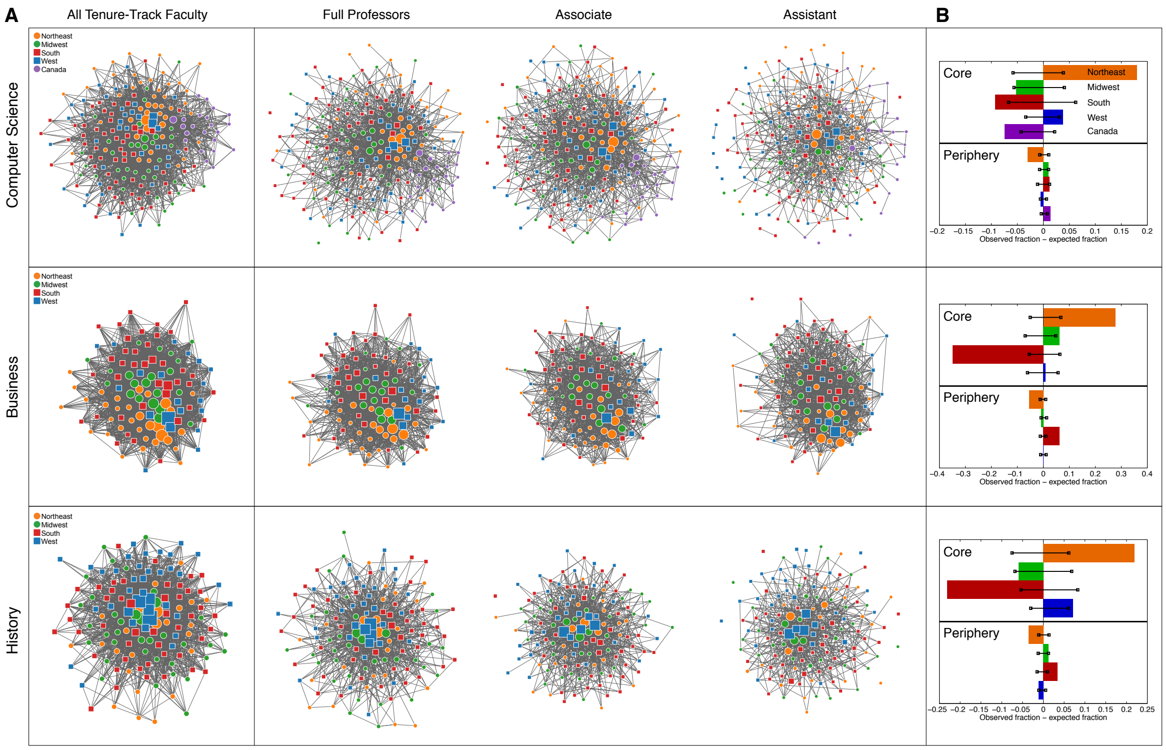

Figure S13: Geographic structure of faculty hiring. (**A**) Network visualizations for faculty hiring, showing all current tenure-track faculty (Full, Associate or Assistant Professors), for computer science, business and history. Vertex size is proportional to the total number of faculty produced (out-degree k_o), and is assigned a color according to its geographic location among the U.S. Census regions or Canada. (**B**) Over-representation of departments by geographic location within the core (top 15%) and periphery, relative to the expected representation by the fraction of all institutions by region; bars are ordered by region in the same way in each panel. Black bars indicate the 25–75% quantiles, by Monte Carlo, indicating a common over-representation of high-prestige institutions in the Northeast and under-representation in the South (and Canada for computer science). Other regions are within expectation.