A Century of Faculty Globalization and University Rise: From US-Born, Foreign-Educated to Foreign-Born, US-Educated

Christian Chacua [*] Frank Neffke [†]

March 27, 2026

Please do not cite or distribute without permission of the authors.
Click here to download latest version

Abstract

U.S. universities transformed from provincial institutions into global research leaders in less than a century. A fundamental component of their success was their ability to attract and employ talented researchers. We examine this transformation by analyzing over 396k biographical entries of U.S. faculty from twelve editions of American Men and Women of Science, tracking faculty hired between 1880 and 1990. Our findings reveal that U.S. universities shifted from relying on U.S.-born, foreign-educated faculty to attracting foreign-born, U.S.-educated faculty. In the early 1900s, approximately 20% of hired faculty were foreign-educated, predominantly German-trained Americans. By 1990, around 30% were foreign-born and increasingly U.S.-educated, with the share of foreign-born, U.S.-educated faculty rising from 7.5% to 20%. This shift coincided with a geographic diversification from German dominance to global representation, including Asia and the Global South, as individual universities evolved from maintaining distinct foreign partnerships to developing dense global networks. Moreover, we observe an association between the presence of foreign-educated faculty and changes in their host institutions' research trajectories, particularly in the late 1800s.

Key words: academic careers, faculty hiring networks, migration of scientists, history of US universities, knowledge diffusion.
JEL: O15, I23, J61, O30, C45

[*]Growth Lab, Center for International Development - Harvard University, Cambridge, MA, United States
[†]Complexity Science Hub, Viena, Austria

American universities transformed from provincial institutions to global leaders in scientific research in less than a century (Geiger, 1986). A fundamental component of their success relied on their capacity to attract and employ talented researchers. We examine the educational backgrounds and career trajectories of multiple generations of leading U.S. scientists to understand the reliance of U.S. academic research institutions on foreign-educated and foreign-born faculty.

We digitized over 396k biographical entries, representing approximately 119k U.S. faculty, listed across twelve editions of the series "American Men and Women of Science" (Cattell, 1906b; Kalte et al., 2005), published between 1906 and 2005. From this dataset, we identify U.S. faculty first hired by American universities between 1880 and 1990, tracking where they acquired their advanced research education and how they influenced their host institutions' research trajectories.

Our findings reveal that over a century, U.S. universities transitioned from relying on *U.S.-born, foreign-educated* to *foreign-born, U.S.-educated* faculty (see Figure 1). Until the 1900s, around 20% of faculty hired in the U.S. were foreign-educated, predominantly German-educated, U.S.-born (see Figure 2a). By 1920, the share of foreign-educated faculty declined to 5% and, although it has increased in recent periods, it did not exceed 13%. Simultaneously, U.S. universities have become increasingly reliant on foreign-born scientists, whose share grew from approximately 10% in 1880 to around 30% by 1990. Notably, these foreign-born faculty were increasingly educated in U.S. institutions, with the share of *foreign-born* among *U.S.-educated* faculty rising from around 7.5% in 1880 to around 20% by 1990.

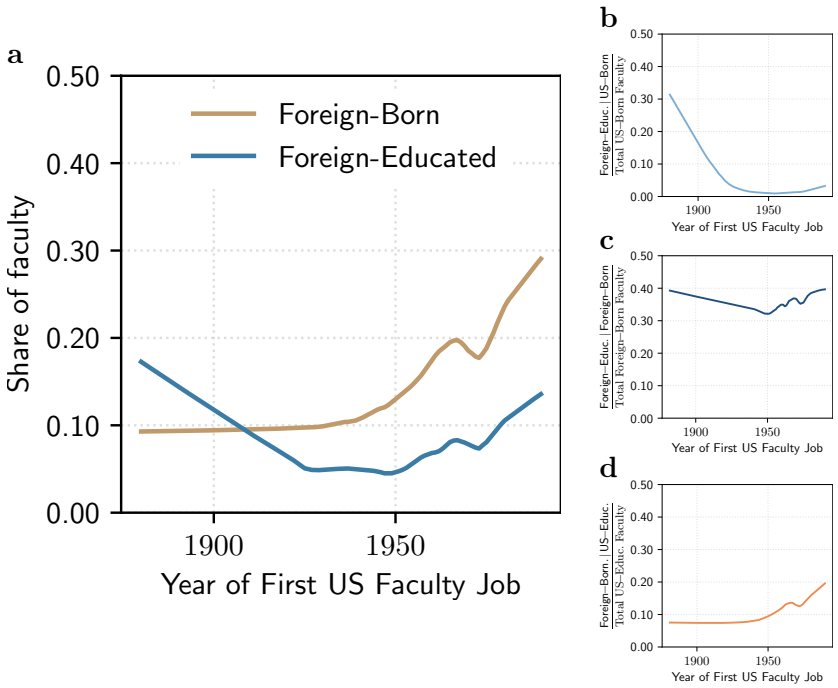
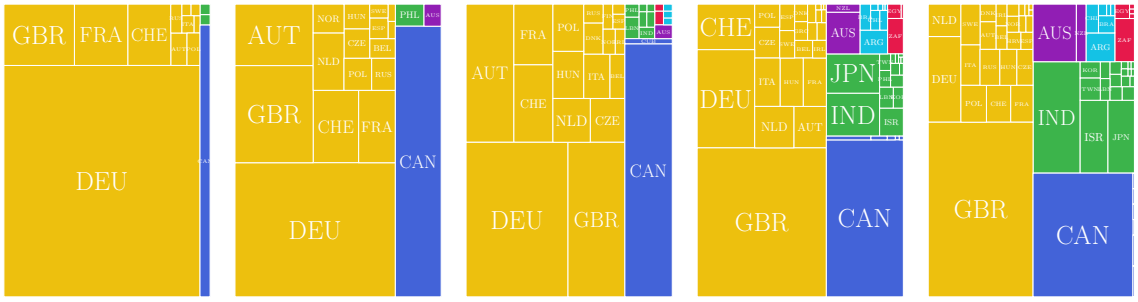


Figure 1: Foreign-Born and Foreign-Educated by Year of First US Faculty Job

Source: Own estimations using data from AMWS. **a**: Share of Foreign-Born and Foreign-Educated among the total faculty. **b**: Foreign-Educated among US-born faculty. **c**: Foreign-Educated among Foreign-born faculty. **d**: Foreign-Born among US-Educated faculty.

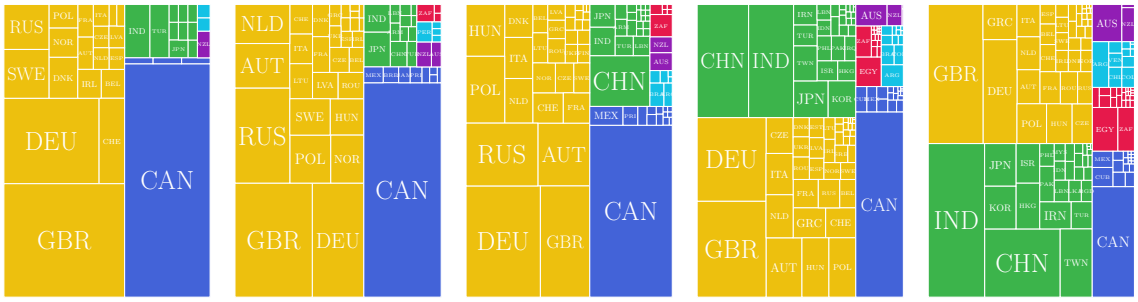
Concerning the geographic sources of foreign education and place of birth, we observe a

general trend for U.S. universities to open to a more diversified and global pool of faculty. The geographic sources of foreign education diversified (see Figure 2a), with the United Kingdom, Canada, and India becoming increasingly important and Germany's influence declining. The origins of foreign-born faculty shifted from European and Canadian dominance in the late 1800s to greater representation from Asia, Europe, and other regions of the Global South (see Figure 2b).



1880-1910 1911-1930 1931-1950 1951-1970 1971-1990

(a) Education places of the foreign-educated



1880-1910 1911-1930 1931-1950 1951-1970 1971-1990

(b) Birthplaces of the foreign-born

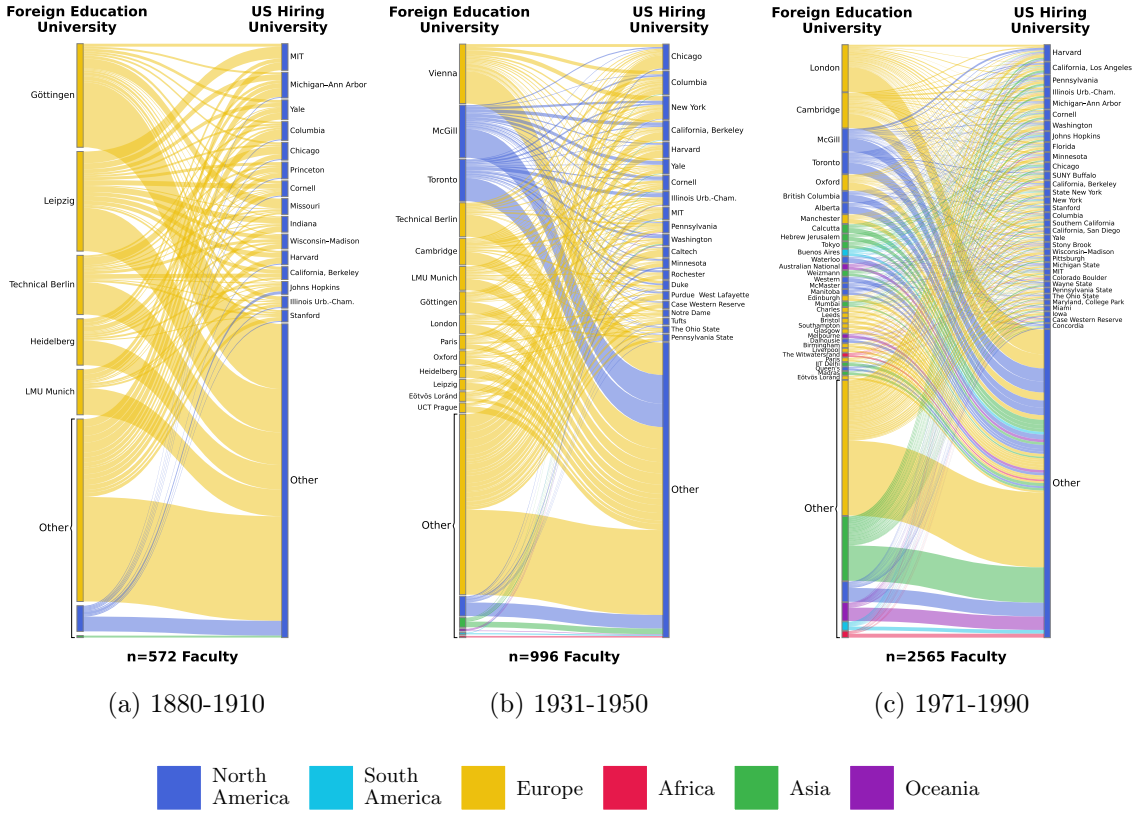
North America | South America | Europe | Africa | Asia | Oceania

Figure 2: Composition of the foreign-educated and the foreign-born

Source: Own estimations using data from AMWS. Year cohorts defined according to the year of first faculty job in the US. Using current country boundaries as defined in Geonames (Wick, 2021) and GADM (Global Administrative Areas, 2022).

This shift was facilitated by U.S. universities' expanding international faculty hiring networks, which provided access to foreign-educated scientists from diverse institutions worldwide. We document how different U.S. universities maintained distinct ties with foreign universities. In 1906, MIT's primary source of foreign faculty was Leipzig, while the University of Missouri recruited heavily from Göttingen (see Figure 3a). In recent periods, these networks have become denser and more diverse, with universities as a group drawing talent globally while maintaining overrepresentation of specific partners (see Figure 3c).

We also examine the long-term impact of different cohorts of foreign-educated faculty on the research output of their host universities many years after their arrival. Specifically, we

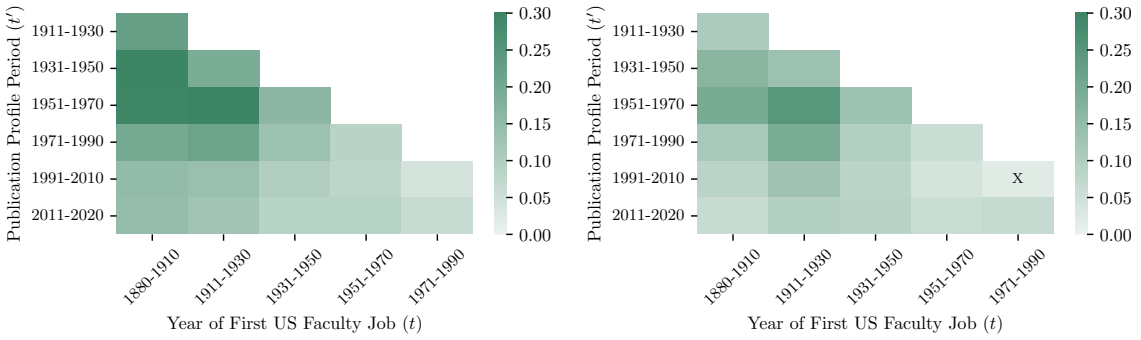


(a) 1880-1910 (b) 1931-1950 (c) 1971-1990

Figure 3: Evolution of Faculty Hiring Networks for Foreign-Educated

Source: Own estimations using data from AMWS. Nodes represent universities, while links illustrate the flow of faculty from a foreign university of education (left) to a hiring US university (right). The panels display the network across three time windows, defined by the year of the first US faculty appointment: **(a)** 1880–1910, **(b)** 1931–1950, and **(c)** 1971–1990. Flows are colored according to the continent of the foreign university of education. Foreign universities (left) are ranked by the total number of graduates, with the bottom 10% aggregated into an "Other" category, split by continent. US universities (right) are ranked by the total number of faculty hired, with the bottom 10% grouped as "Other."

assess whether the influx of scientists from foreign institutions with outstanding track records in a particular academic field is linked to an increase in publications in that field at their host universities (see Figure 4). While the arrival of foreign-educated scientists had, in most periods, a positive effect on their host universities' research output, the impact of *foreign-born foreign-educated* faculty was long-lasting and more persistent for earlier cohorts (1880-1910 and 1911-1930) (see Figure 4, panel a). The effect of U.S.-born, foreign-educated faculty was generally smaller, and only considerable for the same earlier cohorts (see Figure 4, panel b). These results suggest that early cohorts of foreign-educated faculty, particularly those who were foreign-born, had a greater impact when U.S. universities were still developing their research capabilities. Over time, however, their influence diminished as American institutions increasingly relied on domestically-trained faculty.



(a) Foreign-Born (b) US-Born

Figure 4: Impact of Foreign-Educated on the publication profiles of their first hosting US universities

Source: Own estimations using data from AMWS. Note: Cells marked with an "X" indicate results that are not statistically significant. Panel (a) displays how the presence of foreign-born, foreign-educated faculty from institutions outstanding in a particular field relates to the number of research papers produced in that field at US universities. Panel (b) shows similar results for US-born, foreign-educated faculty.

Our work contributes to several literatures. First, it provides quantitative evidence for debates about the emergence of U.S. academic research (Geiger, 1986; Cole, 2010; Kerr, 2001) and the historical formation of American research capabilities (Urquiola, 2020; MacLeod and Urquiola, 2021; Graham and Diamond, 2004; Cole, 2010). Second, it informs ongoing discussions about foreigners' contributions to U.S. science and innovation (Bernstein et al., 2022), particularly regarding migration and knowledge diffusion (Hornung, 2014; Lissoni, 2018; Ganguli, 2015) and migrants as agents of change (Hausmann and Neffke, 2019; Bahar and Rapoport, 2018; Bahar et al., 2020; Miguelez and Morrison, 2022). Third, it adds to research on U.S. faculty origins (Clauset et al., 2015; Burris, 2004; Jones and Sloan, 2020) and the U.S. academic elite (Yuret, 2017,0). Finally, it assembles a comprehensive geocoded dataset tracking leading American scientists' careers over more than a century.

Our findings also inform current geopolitical debates about foreigners' roles in U.S. academic research. Our analysis suggests that when America was not yet a global academic leader, foreign education, primarily of U.S.-born scholars, helped the country acquire scientific capabilities to rise to prominence. Today, the U.S. relies on foreign-born talent to maintain its dominant position in science and innovation. While foreign-born scientists may

not dramatically alter universities' publication profiles, they expand the talent pool and may bring diverse global perspectives. Historically, Germany attracted the best U.S.-born scientists because German institutions led the world; today, America holds that advantageous position. Restricting this talent flow could impede U.S. research progress and directly benefit competitors.

References

Bahar, D., Choudhury, P., and Rapoport, H. (2020). Migrant inventors and the technological advantage of nations. *Research Policy*, 49(9):103947.

Bahar, D. and Rapoport, H. (2018). Migration, knowledge diffusion and the comparative advantage of nations. *The Economic Journal*, 128(612):F273–F305.

Balassa, B. (1965). Trade liberalisation and "revealed" comparative advantage. *The Manchester School*, 33(2):99–123.

Bernstein, S., Diamond, R., Jiranaphawiboon, A., McQuade, T., and Pousada, B. (2022). The contribution of high-skilled immigrants to innovation in the united states. Technical report, National Bureau of Economic Research.

Burris, V. (2004). The academic caste system: Prestige hierarchies in PhD exchange networks. *American sociological review*, 69(2):239–264.

Cattell, J. M. (1906a). A statistical study of American men of science: The selection of a group of one thousand scientific men. *Science*, 24(621):658–665.

Cattell, J. M. (1906b). *American men of science: A biographical directory*, volume 1. The Science Press, New York.

Cattell, J. M. (1910). A further statistical study of American men of science. *Science*, 32(828):672–688.

Cattell, J. M. (1917). Families of American Men of Science. *The Scientific Monthly*, 4(3):248–262.

Clauset, A., Arbesman, S., and Larremore, D. B. (2015). Systematic inequality and hierarchy in faculty hiring networks. *Science advances*, 1(1):e1400005.

Cole, J. R. (2010). *The great American university: Its rise to preeminence, its indispensable national role, why it must be protected*. PublicAffairs.

Di Iasio, V. and Miguelez, E. (2022). The ties that bind and transform: knowledge remittances, relatedness and the direction of technical change. *Journal of Economic Geography*, 22(2):423–448.

Digital Science (2021). Global research identifier database (grid). data retrieved from the 18th version, https://doi.org/10.6084/m9.figshare.c.3812929.v18.

Ganguli, I. (2015). Immigration and ideas: what did Russian scientists "bring" to the United States? *Journal of Labor Economics*, 33(S1):S257–S288.

Geiger, R. L. (1986). *To advance knowledge: The growth of American research universities, 1900-1940*. Oxford University Press.

Global Administrative Areas (2022). Gadm, the database of global administrative areas. https://gadm.org/. Version 4.1, July 2022.

Godin, B. (2007). From eugenics to scientometrics: Galton, Cattell, and men of science. *Social studies of science*, 37(5):691–728.

Graham, H. D. and Diamond, N. (2004). *The rise of American research universities: Elites and challengers in the postwar era*. JHU Press.

Hammarfelt, B., De Rijcke, S., and Wouters, P. (2017). From eminent men to excellent universities: University rankings as calculative devices. *Minerva*, 55(4):391–411.

Hausmann, R. and Neffke, F. (2019). The workforce of pioneer plants: The role of worker mobility in the diffusion of industries. *Research Policy*, 48(3):628–648.

Hidalgo, C. A., Balland, P.-A., Boschma, R., Delgado, M., Feldman, M., Frenken, K., Glaeser, E., He, C., Kogler, D. F., Morrison, A., Neffke, F., Rigby, D., Stern, S., Zheng, S., and Zhu, S. (2018). The principle of relatedness. In Morales, A. J., Gershenson, C., Braha, D., Minai, A. A., and Bar-Yam, Y., editors, *Unifying Themes in Complex Systems IX*, pages 451–457, Cham. Springer International Publishing.

Hornung, E. (2014). Immigration and the diffusion of technology: The Huguenot diaspora in Prussia. *American Economic Review*, 104(1):84–122.

Jones, T. R. and Sloan, A. (2020). Staying at the top: The Ph.D. origins of economics faculty. In *2020 APPAM Fall Research Conference*. APPAM.

Kalte, P. M., Nemeh, K. H., and Schusterbauer, N. (2005). *American Men & Women of Science*, volume 1 A-B. Thomson Gale, 22 edition.

Kerr, C. (2001). *The uses of the university*. Harvard University Press.

Kim, S. D. and Moser, P. (2021). Women in science. lessons from the baby boom. Technical report, National Bureau of Economic Research.

Li, Y. and Neffke, F. (2022). Relatedness in regional development: in search of the right specification. *arXiv preprint arXiv:2205.02942*.

Lissoni, F. (2018). International migration and innovation diffusion: an eclectic survey. *Regional Studies*, 52(5):702–714.

MacLeod, W. B. and Urquiola, M. (2021). Why does the United States have the best research universities? Incentives, resources, and virtuous circles. *Journal of Economic Perspectives*, 35(1):185–206.

Miguelez, E. and Morrison, A. (2022). Migrant inventors as agents of technological change. *The Journal of Technology Transfer*, pages 1–24.

Montani, I., Honnibal, M., Honnibal, M., Landeghem, S. V., Boyd, A., Peters, H., McCann, P. O., Samsonov, M., Geovedi, J., O'Regan, J., Altinok, D., Orosz, G., Kristiansen, S. L., Miranda, L., de Kok, D., Roman, Bot, E., Fiedler, L., Howard, G., Edward, Phatthiyaphaibun, W., Tamura, Y., Bozek, S., murat, Daniels, R., Amery, M., Böing, B., Vanroy, B., and Tippa, P. K. (2022). explosion/spaCy: v3.3.0: Improved speed, new trainable lemmatizer, and pipelines for Finnish, Korean and Swedish.

Moreno-Monroy, A. I., Schiavina, M., and Veneri, P. (2020). Metropolitan areas in the world. delineation and population trends. *Journal of Urban Economics*, page 103242.

Moser, P. and San, S. (2020). Immigration, science, and invention. Lessons from the quota acts. *Working paper*.

Neffke, F. M. H. (2019). The value of complementary co-workers. *Science Advances*, 5(12):eaax3370.

Priem, J., Piwowar, H., and Orr, R. (2022). OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. *arXiv preprint arXiv:2205.01833*.

Schiavina, M., Moreno-Monroy, A., Maffenini, L., and Veneri, P. (2019). GHS-FUA R2019A - GHS functional urban areas, derived from GHS-UCDB R2019A. *European Commission, Joint Research Centre (JRC) [Dataset]*.

Schubert, A. and Braun, T. (1986). Relative indicators and relational charts for comparative assessment of publication output and citation impact. *Scientometrics*, 9(5):281–291.

Trytten, M. and Shapiro, T. R. (1951). The earnings of american men of science. *Science*, 113(2935):345–347.

Urquiola, M. (2020). *Markets, Minds, and Money*. Harvard University Press.

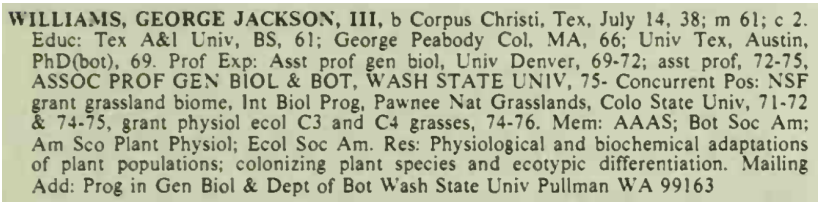
Wick, M. (2021). GeoNames. Free Gazetteer Data retrieved in November 2021. http://download.geonames.org/export/dump/allCountries.zip [Dataset].

Yuret, T. (2017). An analysis of the foreign-educated elite academics in the united states. *Journal of Informetrics*, 11(2):358–370.

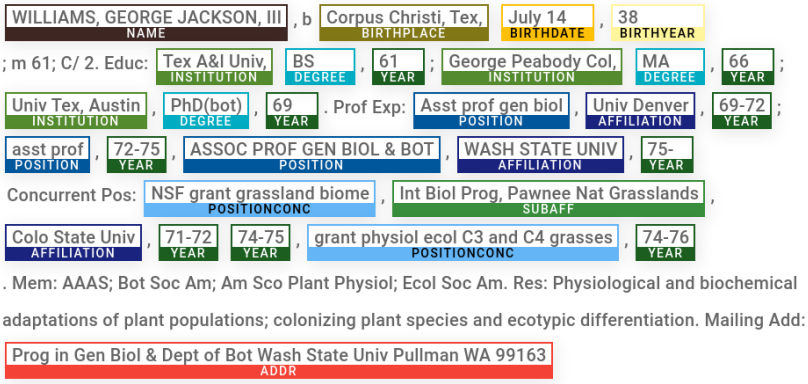
Yuret, T. (2018). Path to success: An analysis of us educated elite academics in the united states. *Scientometrics*, 117(1):105–121.

A Data

Our primary data comes from 396k biographical entries on 119k leading US faculty. Biographical data comes from 12 different releases (1906, 1910, 1921, 1933, 1938, 1944, 1949, 1961, 1976, 1986, 1995, 2005) of the series American Men and Women of Science (AMWS).[1] As a single release can comprise multiple volumes, a total of 39 books and more than 37k pages have been processed. For every page, we have taken digitized images (see panel (a) of figure A1) and transformed them into text that can be read using a set of customized NLP Models for Name Entity Recognition (NER). In particular, we extract information on dates and places of birth, education, and work (see panel b of figure A1). More technical details of this procedure can be consulted in the appendix C.



(A) An AMWS raw biography



(B) A labeled biography

Figure A1: Sample Biography

Source: Own labeling using data from AMWS. Note: Panel (A) shows a raw biography, as it is extracted from the original books; Panel (B) shows a labeled biography after applying our NLP-NER models to extract all the relevant textual entities.

By leading American scientists, we follow the definition adopted by AMWS. That is, scientists who have conducted "a significant proportion of their work [...] in North America" (Kalte et al., 2005, p. VII). Therefore, foreign-born scientists are included in our analysis. Moreover, the scientists listed in the books are meant to have made significant contributions to their fields. Peers usually nominate them, and the significance of their contributions is assessed according to three criteria: distinguished achievements, quality of research, and positions attained. AMWS is not intended to be a census of North American scientists but a sample of the people who have driven science's development in North America.

[1] Although first book releases were titled "American Men of Science," they also included women (Kim and Moser, 2021). The current title, "American Men and Women of Science" was adopted in 1971.

AMWS results from James McKeen Cattell's psychological works on the characteristics of the "American Men of Science" (Cattell, 1906a,9,9,9). Although Cattell was a controversial figure for being in favor of eugenics (Godin, 2007; Hammarfelt et al., 2017), the information he collected has been precious for prosopographical and historical studies (Trytten and Shapiro, 1951). More recently, AMWS has become a regular publication that involves multiple advisory board members. Moreover, economists have started using AMWS data, as some historical periods facilitate the design of robust empirical strategies. For example, Moser and San (2020) study the effects of migration policies in invention during the migration quota acts, and Kim and Moser (2021) analyze the impact of children on the careers of women scientists during the baby boom.

As our focus is on the US faculty, we only consider the scientists who have worked as professors the US at any time of their careers. We exclude visiting professors and emeritus professors from our analyses, as our focus is on tenure-track professors. To keep historical consistency, we only focus on the hires by continental US universities, excluding scientists who started their careers in Alaska and Hawaii. We group faculty in time cohorts, deined according to the year of first position as a professor.

To locate universities, we matched universities of education and universities of work with the June 2021 release of the GRID database (Digital Science, 2021). In addition, as some institutions have changed their names or merged with other institutions, we manually corrected the top 300 institutions in every book released, corroborating data with alternative sources such as Wikipedia and the universities' websites. We rely on aggregated data at the city level in some of the indicators used in our econometric estimations (see section B for more details on our econometric estimations and variable definitions). To define the boundaries of a city, we follow the definition of Functional Urban Areas (FUA) established by the Global Human Settlement Layer (GHSL) project (Moreno-Monroy et al., 2020; Schiavina et al., 2019).

To determine the places of birth for faculty members, we geocoded the locations provided in their records. We began by cleaning the input text, removing non-ASCII characters that often corresponded to OCR errors, and by identifying each subcomponent in the birthplace string. For instance, some entries specify a "city, US state" pair, others use "city, country," and some list only the city or the country. Each place of birth was matched to entries in the Geonames database, with the goal of preserving hierarchical relationships and achieving geocoding at the finest level of detail available. In cases where a match was not found in Geonames, we used the raw string to query the Google Maps API. A faculty member was classified as foreign-born if the reported birthplace was outside the current boundaries of the United States, including Alaska and Hawaii. We adopted modern country boundaries as defined in Geonames (Wick, 2021), when a match is available, and GADM (Global Administrative Areas, 2022), when geocoded using Google Maps API. This is particularly important since many European boundaries have changed during our period of analysis. This approach is feasible because most faculty members specify city names in the place of birth field.

In our econometric estimations, we also obtain data from the January 2024 snapshot of OpenAlex (Priem et al., 2022). We use data on the documents with an attributed academic field (Concept), published between 1880 and 2020, to compute 1) the publication profile and the scientific specialization of the considered US research universities; 2) the publication profile and the scientific specialization of the foreign cities of education; and 3) the knowledge

proximity among academic fields, computing the co-occurrences of academic fields in OpenAlex works. In all our analyses, we consider works as our unit of observation, and consider all work types listed in OpenAlex, as long as they have an associated academic field at level 1 (284 concepts). This choice is justified, as historical works do not always have an assigned work type, or they are not correctly classified.

B Empirical strategy

We estimate regression models to analyze how foreign education has affected the publication profiles of the hosting US universities of work. In particular, we examine the relationship between the fractional publication count of a US university u in academic field f during a more recent period t', transformed as $\text{asinh}(y_{u,f,t'})$, and the composition of its foreign-educated entry-level faculty hired during a historical period t, such that $t < t'$.[2] Our estimations cannot be read as causal effects, as it is not our objective to account for the numerous historical events and endogenous factors that may have affected the performance of US universities.

Equation (1) shows the baseline model specification. We model $\text{asinh}(y_{u,f,t'})$ using as main independent variables: our migration measure $\text{asinh}(\boldsymbol{m}_{u,f,t})$; a mean reversion term $\text{asinh}(y_{u,f,t})$, based on the fractional publication count of university u in field f during historical period t; the field density term $\text{asinh}(w_{u,f,t})$, measuring the relatedness of field f to the existing research portfolio of university u at time t; a vector of university fixed effects δ_u; and a vector of academic field fixed effects δ_f.

$$\text{asinh}(y_{u,f,t'}) = \beta_0 + \beta_1 \,\text{asinh}(\boldsymbol{m}_{u,f,t}) + \beta_2 \,\text{asinh}(y_{u,f,t}) + \beta_3 \,\text{asinh}(w_{u,f,t}) + \delta_u + \delta_f + \epsilon \quad (1)$$

All continuous variables entering the model, including the dependent variable, are transformed using the inverse hyperbolic sine $\text{asinh}(x)$ prior to estimation. The asinh transformation behaves approximately linearly around zero and approximately logarithmically for large x, while being defined at zero and for all nonnegative values without the need for adding a constant. This property preserves zero-valued observations in the estimation sample, which is particularly important for the migration variable and for the outcome: a university may have received foreign-educated faculty from cities specialized in a given field without having previously published in that field, or may have begun publishing in a new field precisely as a result of such migration flows. Discarding zero-valued observations would introduce sample selection bias against the very mechanism the model aims to identify.

Our own-field f migration variable, $m_{u,f,t}$ (equation (2)), measures the direct effect of a cohort of foreign-educated entry-level faculty on the subsequent research trajectory of a university. It relies on the idea that historical specialization patterns at foreign education cities may drive current specialization at US universities—for example, a university that is

[2] We use fractional publication counts rather than integer counts or citations for two reasons. First, papers in OpenAlex are assigned to multiple academic concepts simultaneously. Under fractional counting, a paper assigned to k concepts contributes $1/k$ to each concept's total, ensuring that the aggregate publication count across all concepts equals the total number of published papers. Second, citation data in OpenAlex are limited for papers published in the early twentieth century, making fractional publication counts a more reliable indicator for historical periods.

currently strong in chemistry may have been nourished by foreign-educated faculty from cities that were historically specialized in chemistry. The variable is defined as a weighted count: for each university u and historical period t, we sum across all foreign education cities o the number of scientists $l_{o,u,t}$ who trained in city o and worked at u during t, weighting each city by a binary indicator of whether it held a comparative advantage in field f during t.[3] Formally, $RCA^{\text{binary}}_{o,f,t}$ takes the value of one if $RCA_{o,f,t} > 1$ and zero otherwise. $m_{u,f,t}$ is expressed as a raw count of scientists, so that its scale reflects the absolute size of the relevant migration flows received by each institution.

$$m_{u,f,t} = \sum_{o=1}^{O} l_{o,u,t} \times RCA^{\text{binary}}_{o,f,t} \tag{2}$$

Concerning our Revealed Comparative Advantage indexes $RCA_{o,f,t}$ used in equation (2) and all our relative measures, we follow a definition analogous to Balassa (1965) and Schubert and Braun (1986). The index $RCA_{o,f,t}$ (equation (3)) identifies foreign education cities o that were disproportionately specialized in producing publications in field f relative to the expected share given their overall publication output. All publication counts entering the computation of $RCA_{o,f,t}$ are based on fractional counting from OpenAlex. We compute these measures at the city level, rather than at the institutional level, to reduce biases arising from poor geocoding and disambiguation of non-US affiliations in historical data, and to circumvent the multiple changes in university names and administrative structures that are difficult to trace across long time horizons.[4] The index is binarized: $RCA^{\text{binary}}_{o,f,t}$ equals one if $RCA_{o,f,t} > 1$, and zero otherwise.

$$RCA_{o,f,t} = \frac{\text{Observed share of city } o \text{ in total fractional output of field } f \text{ at time } t}{\text{Expected share of city } o \text{ among all cities at time } t} \tag{3}$$

We also control for the direct academic field density $w_{u,f,t}$ of US university u around field f at time t, as shown in equation (4). Following the principle of relatedness (Hidalgo et al., 2018), universities are expected to develop strength in fields related to those they have already mastered. Accordingly, $w_{u,f,t}$ accounts for how accessible it was for university u to expand into field f, given the fields in which it held a comparative advantage during historical period t. The variable takes the form of a weighted sum, in which we add the normalized pairwise concept proximity $\phi_{ff'}/\phi_f$ of each surrounding field f' around f, with nonzero weight only if university u held a comparative advantage in field f' in t. We compute the concept proximity matrix from the co-occurrence of academic concepts across all articles in OpenAlex, covering the full observation window from 1880 to 2020, which allows us to capture relationships between established and emerging fields in a time-consistent way.[5] Our proximity measures $\phi_{ff'}$ and $\phi_f = \sum_{f' \neq f} \phi_{ff'}$ are therefore time-independent.[6] As in equation (2), we binarize

[3] Similar migration variables multiplied by a measure of comparative advantage can be found in Bahar and Rapoport (2018) and in Di Iasio and Miguelez (2022).

[4] We have cleaned errors related to the institutional affiliations of scientists in our AMWS sample, but further work is required to link historical publications in OpenAlex to current institutional identifiers.

[5] Academic concepts in OpenAlex are assigned using machine learning methods, which reduces problems associated with differences in field classification schemes over time.

[6] For our relatedness matrices we rely on the RCA transformation proposed by Li and Neffke (2022):

the revealed comparative advantage index $RCA_{u,f',t}$, and all fractional publication counts entering its computation are drawn from OpenAlex.

$$w_{u,f,t} = \sum_{f' \neq f} \frac{\phi_{ff'}}{\phi_f} \times RCA^{\text{binary}}_{u,f',t} \tag{4}$$

The variables t and t' represent specific time intervals rather than individual years. The historical periods t are defined as 1880–1910 (thirty years), 1911–1930, 1931–1950, 1951–1970, and 1971–1990 (twenty years each). The first period spans thirty years to ensure a sufficient sample size for the early decades of the US academic system, during which the number of universities hiring foreign-educated scientists was limited. The recent outcome periods t' are twenty-year windows aligned to the same decade boundaries: 1911–1930, 1931–1950, 1951–1970, 1971–1990, and 1991–2010. We additionally include a ten-year boundary window, 2011–2020, reflecting the right edge of the available data.[7] For each historical period t, we include all valid outcome windows t' satisfying $t' > t$, yielding a stacked panel in which the same historical predictors may be paired with multiple outcome windows. In each historical period, we include only those universities that hired at least one scientist during that interval, allowing the composition of the sample to expand over time in line with the historical growth of the US academic system.

Regarding the estimation method, we estimate OLS regression models with two-way fixed effects. US hiring university fixed effects δ_u absorb time-invariant heterogeneity in university size, prestige, and research orientation, including differences in the overall scale of migration inflows received by each institution. Academic field fixed effects δ_f control for differences in the overall level of publication activity and long-run growth trends across fields. We cluster standard errors by university and academic field to account for potential group-level correlation within institutions and within fields. The estimation sample consists of university–field observations for which the host university had a positive total fractional publication count in the historical period t. All continuous variables are transformed using $\text{asinh}(x)$ prior to estimation, yielding coefficients that can be interpreted analogously to semi-elasticities in a log-linear specification for larger values of x, while maintaining a well-defined transformation at zero.

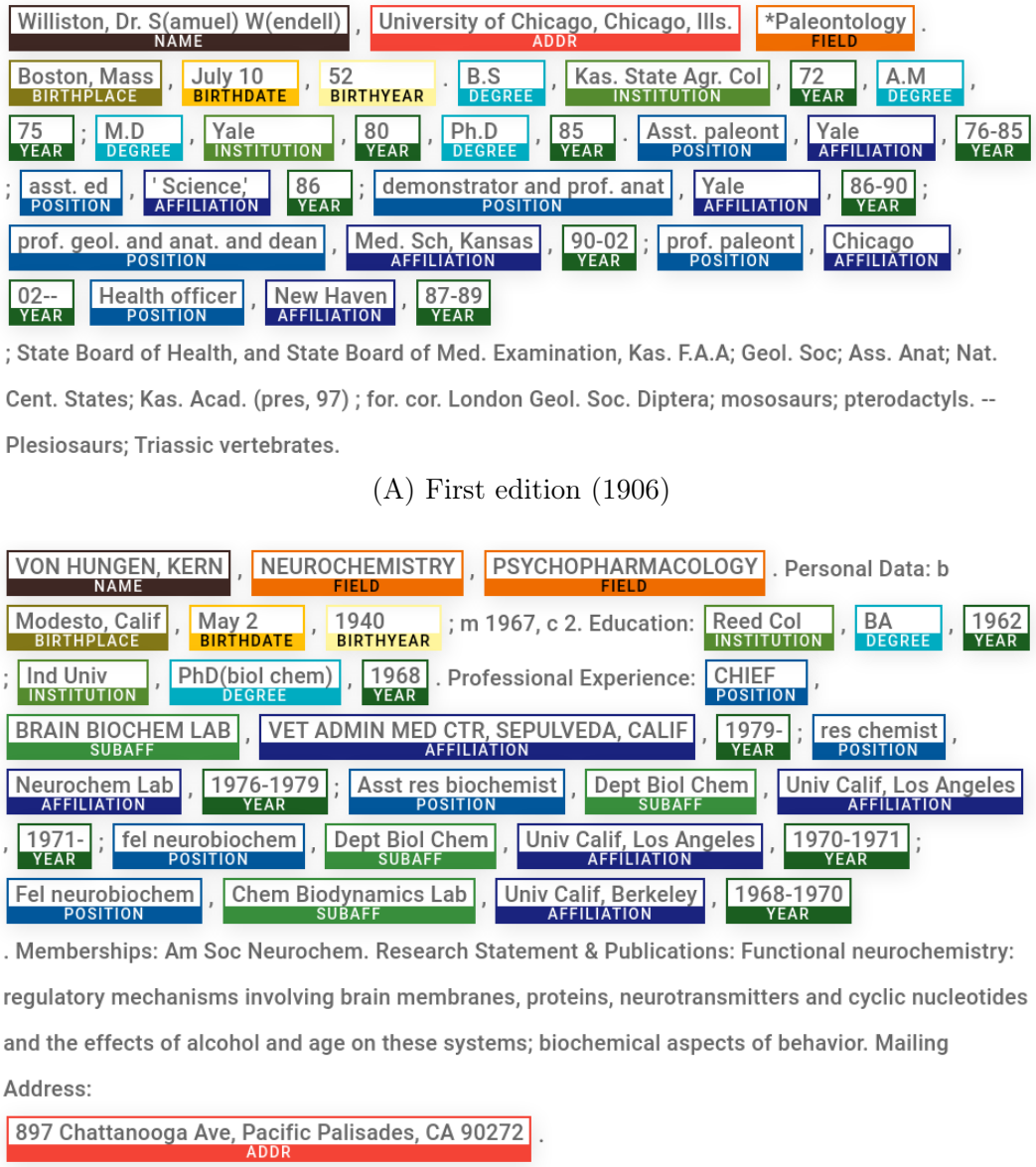
C Sample definition

A typical AMWS biographical entry contains a fixed structure that we parsed using a set of customized Natural Language Processing (NLP) models. As there are some changes in the structure of biographies of the different editions of AMWS (see figure A2), we train five different Name Entity Recognition (NER) models (m0107, m0810, m1316, m19, m22) to extract 18 types of entity labels (see table 1). Every NER model has been trained using a set of 300 to 600 manually labeled biographies and transformer-based models on Python's

$\phi_{ij}^{RCA} = \frac{c_{ij}}{(\sum_i c_{ij})(\sum_j c_{ij})} \times (\sum_i \sum_j c_{ij})$. We additionally apply the monotone transformation $\phi_{ij}^{RCAN} = \frac{\phi_{ij}^{RCA}}{\phi_{ij}^{RCA}+1}$ as in Neffke (2019).

[7] The last outcome window spans ten years rather than twenty due to data availability. To account for differences in window length, we report robustness checks in which all publication counts are normalized to annual rates by dividing by the number of years in each window.

library Spacy 3.0 (Montani et al., 2022).



(A) First edition (1906)

(B) Twenty-second edition (2005)

Figure A2: Sample Biographies

Source: Own labelling using data from AMWS. Note: Panel (A) contains a sample biography for the first release (1906) and panel (B) for the latest considered release (2005).

We apply the trained models to every biographical entry according to its associated book edition. As shown in table 1, the total number of entries, with at least one name entity, is equal to 752,682. Not all the biographies contain the same information, as some are incomplete, and others only reference previous books. For this work, we only focus on biographies, including information on names and at least two more labels, among four possible options (birth year, birthplace, institutions, and affiliations).

As scientists can appear in multiple editions, we match them over time to create a unique scientist identifier. Therefore, we perform fuzzy matching on names, using PostgreSQL

Entity Label	m0107	m0810	m1316	m19	m22	Total
NAME	100003	153673	242491	124032	132483	752682
YEAR	99639	149163	224687	118278	125083	716850
AFFILIATION	98937	147712	223571	117903	123847	711970
ADDR	95709	141403	221965	117291	127909	704277
POSITION	99398	147957	224028	107556	124402	703341
INSTITUTION	97591	144024	223329	117465	118506	700915
FIELD	96480	144988	229883	119994	107218	698563
DEGREE	96881	143731	222948	117221	117508	698289
BIRTHPLACE	95502	139959	219072	114107	109299	677939
BIRTHYEAR	95583	141561	215672	113096	109391	675303
BIRTHDATE	94313	141839	217209	112770	106763	672894
SUBAFF	37727	85460	113112	62838	75685	374822
POSITIONCONC	0	0	128927	76190	0	205117
POSITIONCURR	0	0	0	117704	0	117704
EMAIL	0	0	0	0	64473	64473
CITIZEN	0	10817	21353	12113	10260	54543
DEAD	0	2311	0	0	3292	5603
SUBINS	402	0	0	226	0	628

Table 1: Entities extracted from biography entries

Source: Own estimations using data from AMWS. Note: Counting unique biography entries with at least one entity in each category label.

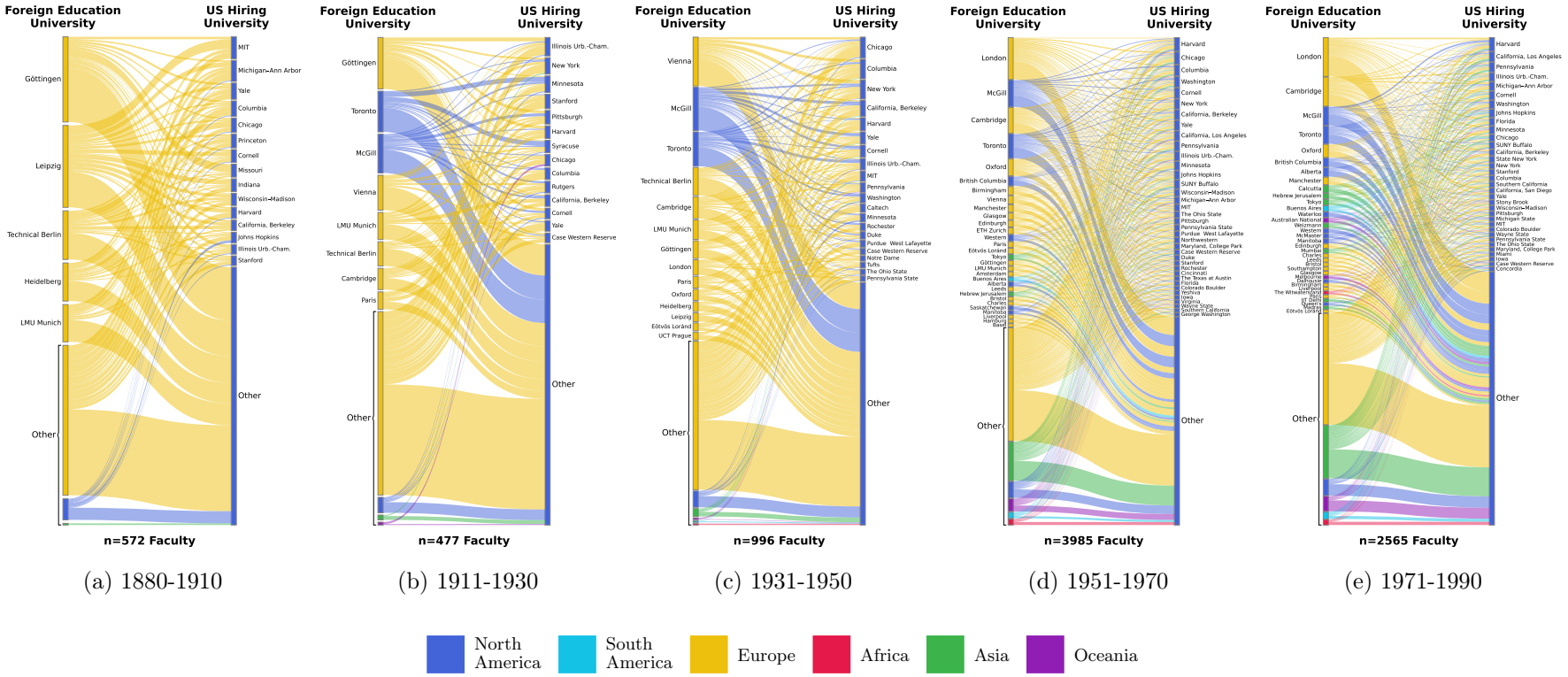
pg_trgm's trigram similarity function and a perfect matching on the birth year. Finally, we set the value of similarity of cleaned names (after removing titles, parenthesis, and giving more weight to first letters) to 0.8. After this process, we get a total of 267,614 unique scientists. This unique scientist identifier is used when counting scientists. Moreover, it is used to extract information by grouping all the available biographical entries associated to an individual.

From the total of 267,614 unique scientists, we identified only those entries corresponding to individuals who hold at least one faculty position in the continental US, as described in the Data section (Appendix A). We excluded from our sample corporate researchers and researchers working in national labs who have not held any faculty positions, as they are out of the scope of this paper and require more extensive disambiguation efforts. Our final sample comprises 396k biographical entries on 119k leading US faculty.

Acknowledgements. We are grateful to Ernest Miguelez, Francesco Lissoni, Andrea Morrison, Ricardo Hausmann, and Muhammed A. Yildirim for their useful remarks. We also received valuable suggestions from participants at the September 9, 2025 NBER Productivity Seminar; the November 24, 2023 Complexity Science Hub Talk; the 3rd Research Policy Online Conference for Early Career Researchers; the 2022 International PhD course on Economic Geography; DRUID 2022; and GeoInno 2022. A first version of this paper was released as "The Changing Geography of American Academic Careers Throughout the 20th Century" and is part of Christian Chacua's doctoral dissertation. However, several methodological changes have been made since then. We acknowledge

the financial support from the French National Research Agency under the program Investissements d'avenir (ANR-10-IDEX-03-02) and the TKC project (ANR-17-CE26-0016). Computer time for this paper was provided by the computing facilities MCIA (Mésocentre de Calcul Intensif Aquitain) of the University of Bordeaux and of the University of Pau and the Adour Region, and by the FASRC Cannon cluster supported by the FAS Division of Science Research Computing Group at Harvard University.

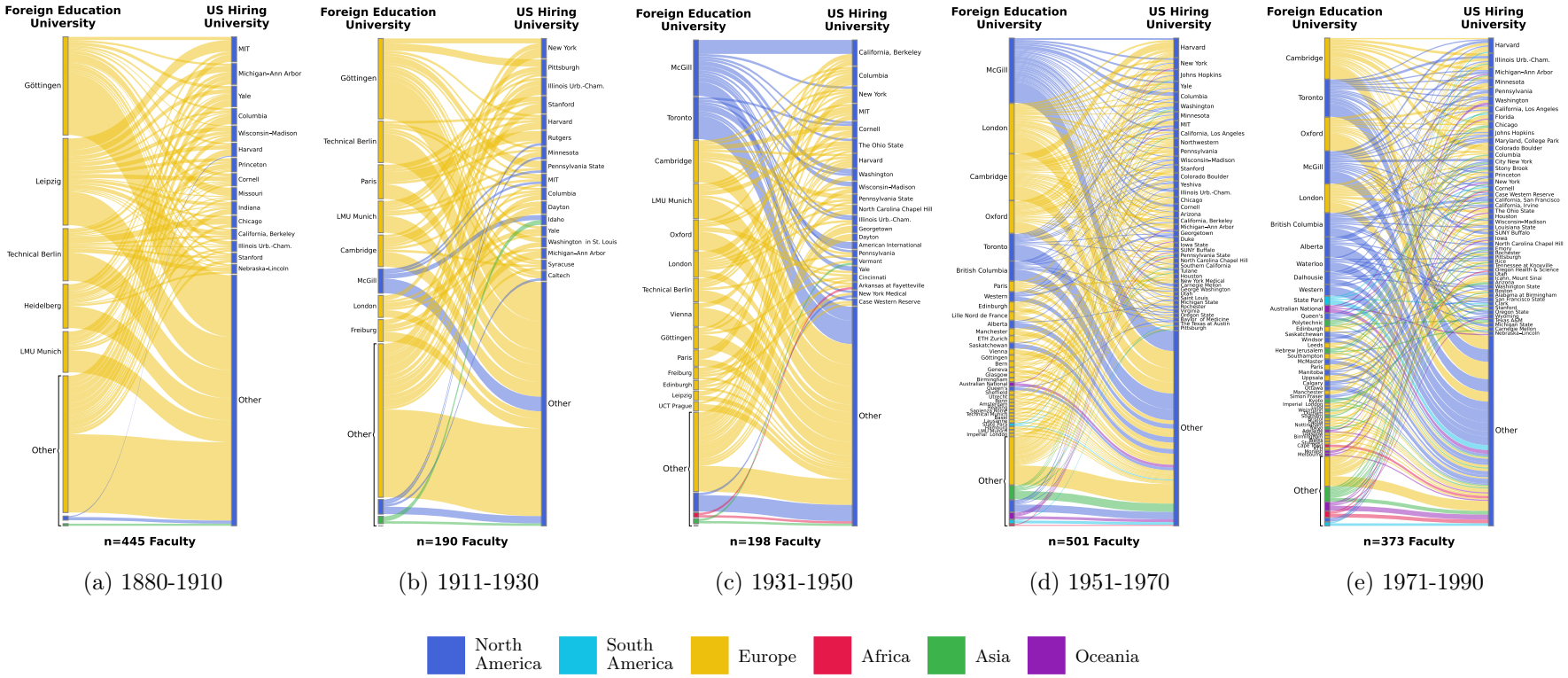
D Additional figures



(a) 1880-1910 (b) 1911-1930 (c) 1931-1950 (d) 1951-1970 (e) 1971-1990

Figure A3: Evolution of Faculty Hiring Networks for all Foreign-Educated

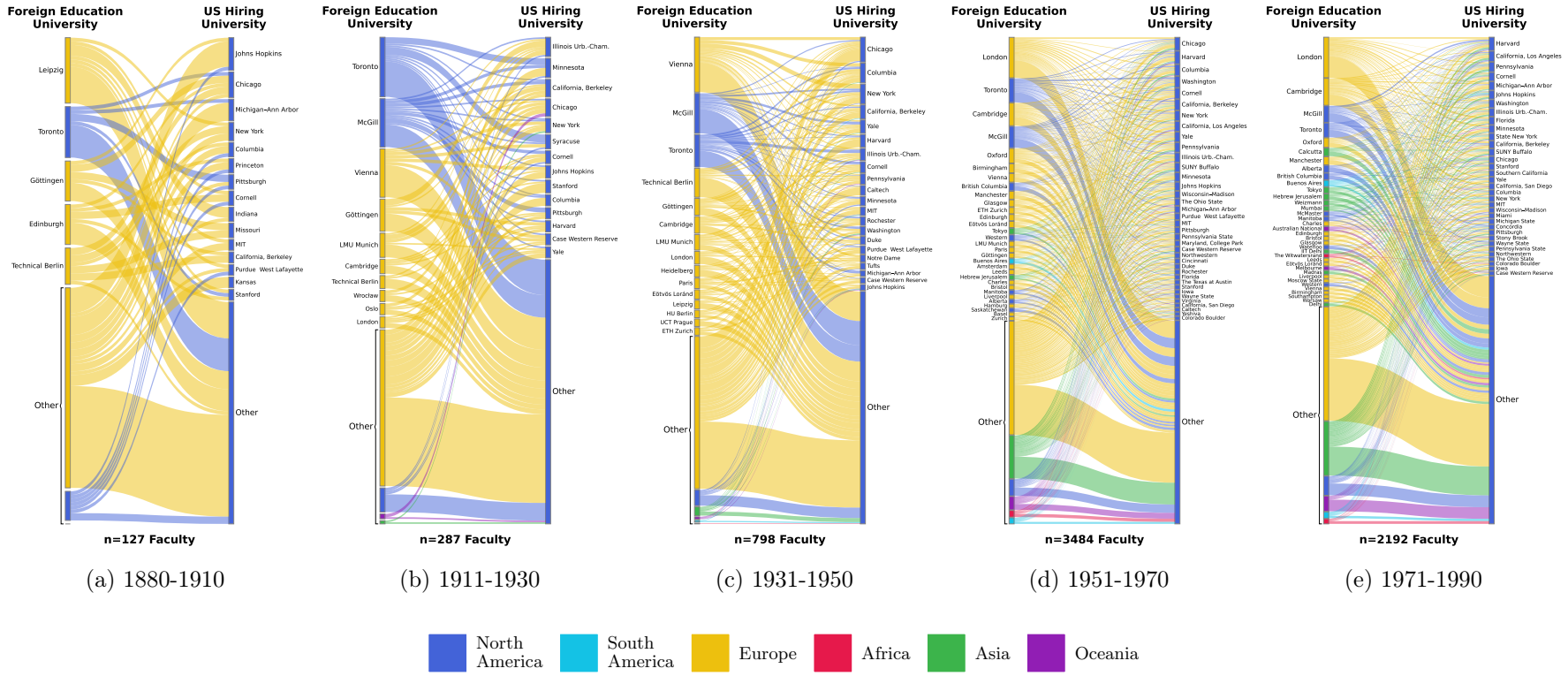
Source: Own estimations using data from AMWS. Nodes represent universities, while links illustrate the flow of faculty from a foreign university of education (left) to a hiring US university (right). The panels display the network across five time windows, defined by the year of the first US faculty appointment. Flows are colored according to the continent of the foreign university of education. Foreign universities (left) are ranked by the total number of graduates, with the bottom 10% aggregated into an "Other" category, split by continent. US universities (right) are ranked by the total number of faculty hired, with the bottom 10% grouped as "Other."



(a) 1880-1910 (b) 1911-1930 (c) 1931-1950 (d) 1951-1970 (e) 1971-1990

Figure A4: Evolution of Faculty Hiring Networks for US-Born Foreign-Educated

Source: Own estimations using data from AMWS. Nodes represent universities, while links illustrate the flow of faculty from a foreign university of education (left) to a hiring US university (right). The panels display the network across five time windows, defined by the year of the first US faculty appointment. Flows are colored according to the continent of the foreign university of education. Foreign universities (left) are ranked by the total number of graduates, with the bottom 10% aggregated into an "Other" category, split by continent. US universities (right) are ranked by the total number of faculty hired, with the bottom 10% grouped as "Other."



(a) 1880-1910 (b) 1911-1930 (c) 1931-1950 (d) 1951-1970 (e) 1971-1990

Figure A5: Evolution of Faculty Hiring Networks for Foreign-Born Foreign-Educated

Source: Own estimations using data from AMWS. Nodes represent universities, while links illustrate the flow of faculty from a foreign university of education (left) to a hiring US university (right). The panels display the network across three five windows, defined by the year of the first US faculty appointment. Flows are colored according to the continent of the foreign university of education. Foreign universities (left) are ranked by the total number of graduates, with the bottom 10% aggregated into an "Other" category, split by continent. US universities (right) are ranked by the total number of faculty hired, with the bottom 10% grouped as "Other."