NBER WORKING PAPER SERIES

CREATING AND CONNECTING US AND CHINA SCIENCE: CHINESE DIASPORA AND RETURNEE RESEARCHERS

Qingnan Xie
Richard B. Freeman

Working Paper 31306
http://www.nber.org/papers/w31306

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2023

We thank Dr Xi Hu from University of Oxford for her support in collecting data from Scopus. We benefited from comments on this and earlier versions of the research from participants in the China Economy Seminar at Harvard University's Department of Economics, the Economics of Science & Engineering Seminar at the Harvard Business School, and the Asia-Pacific Studies Institute Seminar at Duke University. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Creating and Connecting US and China Science: Chinese Diaspora and Returnee Researchers
Qingnan Xie and Richard B. Freeman
NBER Working Paper No. 31306
June 2023
JEL No. F1,F22,F63,I2,I20,J2,J20,J3,J30,O3,O30

ABSTRACT

The close connection between US and China in scientific research and education in the 2000s produced a large group of China-born researchers who work in the US ("diaspora") and a larger group of China-born researchers who gained US-research experience and returned to do their research in China ("returnee"). Analyzing 2018 Scopus data on research papers, we estimate that diaspora researchers contributed to 27% of US addressed papers, and that returnee researchers contributed to 38% of China addressed papers. Both the number of papers with diaspora authors and the number of papers with returnee authors far exceeded the usual measure of US-China collaborative work, papers with both US and China addresses. In terms of quality or impact, papers with diaspora or returnee authors averaged more citations and had higher proportions of publication in high CiteScore journals than other US-addressed or China-addressed papers. Finally, papers with diaspora and/or returnee authors were at the center of the US-China coauthor network and major conduits of research findings between the countries in the network of scientific citations. The benefits of the US-China research connection notwithstanding, the link between the countries' research began to fray from 2018 through the early 2020s, with potential deleterious effects on each country's future research output and on global science writ large to which US and China are the two biggest contributors.

Qingnan Xie
Center for Labor and a Just Economy
Harvard Law School
qingnanx@gmail.com

Richard B. Freeman
NBER
1050 Massachusetts Avenue
Cambridge, MA 02138
freeman@nber.org

In the first two decades of the 21st century, a huge flow of China-born researchers moved from China to the US for graduate education and post-doctorate work (*diaspora researchers*). Many of these researchers returned to China after their US scientific experience to conduct research in China (*returnee researchers*). Using data on 2018 English language journal articles in the Scopus database in 21 physical sciences, engineering, and mathematics fields[1], we analyze the contribution of diaspora and returnee researchers in the creation and dissemination of scientific knowledge in the US and China and document the weakening of that connection from 2018 through the early 2020s.

We present our analysis in three stages. Section one measures the proportion of US and China papers with diaspora or returnee authors and the quality/impact of those papers evinced in citations and the CiteScores[2] of the journal of publication. Section two shows that diaspora or returnee authors were authors on the vast majority of US-China collaborative papers; were links between US and China science throughout their careers; and connected US and China research in the network of citations. Section three examines the late 2010s/early 2020s weakening of the US-China research connection as political and economic tensions rose between the two countries and the COVID-19 pandemic reduced the student and researcher flows which built the connection.

1. Measuring Diaspora/Returnee Research

We identify *diaspora* authors as Chinese-named authors writing at a US address that indicates birth in mainland China from bibliometric data in the 2018 Scopus database[3]. Following Huang & Freeman (2015) and Lin & Chang (2022), we determine the Chinese ethnicity of authors by whether their last names are common Chinese last names in the Chinese Ministry of Public Security's list of Chinese last names[4]. We further differentiate those likely to be mainland born from those likely born in some other location such as Singapore, Taiwan, or Malaysia by whether their first names follow the grammar of mainland China's Hanyu Pinyin translation system[5]. Because the Scopus online system for downloading files provides only an initial for author first names, we obtain full names by randomly sampling 8,000 papers by ethnic Chinese authors[6] via the Scopus API portal (Appendix Table A1, "Main Diaspora Sample"). Our scheme labels *Jianguo Xie* as mainland China born and *John Xie* as non-mainland-China born and labels someone with

[1] Multidisciplinary; Agricultural and Biological Sciences; Biochemistry, Genetics and Molecular Biology; Chemical Engineering; Chemistry; Computer Science; Earth and Planetary Sciences; Energy; Engineering; Environmental Science; Immunology and Microbiology; Materials Science; Mathematics; Medicine; Neuroscience; Nursing; Pharmacology, Toxicology and Pharmaceutics; Physics and Astronomy; Veterinary; Dentistry; Health Professions.

[2] CiteScore is the journal impact indicator calculated by Scopus database defined as the ratio of citations to the journal from scientific documents in the Scopus database over a 4-year period relative to the number of published articles.

[3] The Scopus online downloading files provides the EID of papers, the unique Author ID that differentiates authors with similar names and the last name and first name initial of authors and their addresses on papers. We used the EID and Author ID to collect data on publications and authors through the Scopus API portal.

[4] The 2019 National Name Report (二〇一九年全国姓名报告) lists the most common Chinese last names. It covers the Chinese last names of 84.8% of the mainland population.

[5] Our program distinguishing Chinese first names is available at GitHub: https://github.com/qingnanxie/Chinese-first-name.

[6] The ethnic Chinese authors (Chinese last named authors) in the 8000 samples could have a US address, a China address, or a rest-of-world (non-US & non-China) address, see Appendix Figure A1 for details.

the mainland Hanyu Pinyin translation *Xie* as mainland born as opposed to someone with a non-mainland Pinyin translation *Tes*. In our analysis diaspora authors are those with typical mainland last and first names at a US address in 2018.[7] We define *diaspora papers* as papers with at least one diaspora author. Appendix Figure A1 details how we used a mixture of randomly sampled data with population data to estimate the diaspora share of US research.

To identify *returnee* authors -- China-born researchers writing scientific papers at a China address **after** publishing at least one US-addressed paper -- we searched English language journal articles in the 2018 Scopus database for papers with at least one China address and authors whose names identified them as China-born[8]. We randomly sampled 8,000 of those papers, and then used the Scopus Author ID[9] to retrieve data on *all* of their Scopus-indexed publications, expanding the data to 1.9 million publications (Appendix Table A1 and A2, "Main Returnee Sample"). We identify 2018 *returnee authors* as China-addressed authors with at least one pre-2018 publication where the author had a US-address and define papers with at least one returnee author as *returnee papers*. As this identification excludes China-born researchers who studied or visited the US without having written a US-addressed paper, it is a lower bound on China-addressed authors with some US research experience. Appendix Figure A2 details how we combined the random sample data with population data to estimate the returnee share of US papers.

Our analysis of the country/area where authors conducted their research divides papers into seven mutually exclusive groups based on the addresses of all authors on the paper: US addresses only (USO); China addresses only (CO); US and China addresses only (US-C); US and Rest-of-World (ROW) addresses only (US-ROW); China and ROW addresses only (C-ROW); US, China, and ROW addresses (US-C-ROW); and ROW addresses with no US or China address (ROW). Most US-addressed papers are USO and most China-addressed papers are CO.

Figure 1 displays our measures of the proportion of US-addressed papers with one or more diaspora authors, the proportion of China-addressed papers with one or more returnee authors, and the proportion of papers with at least one author at a US address and at least one author at a China address -- the standard measures of US-China collaboration.

Panel A shows the fact that sparked our work: the huge share of US-addressed papers with diaspora authors compared to the standard metric for US and China collaborative work – papers

[7] This includes US-addressed authors who may also have a China or other country address. Our methodology misses the likely small number of China-born authors who changed their names into non-Chinese names, Chinese ethnicity authors born outside of China but given a Chinese name, and authors with rare Chinese names.

[8] For simplicity we assume that 100% of China-addressed authors are Chinese-named authors. This ignores non-Chinese named authors on those papers. In our sample, 1.8% of China addressed authors did not have a typical Chinese first and last name. Randomly sampling 50 of the authors with non-typical Chinese names, we found 58% had non-Chinese names, 28% were rare Chinese names, 10% were non-mainland Chinese names, 2% were Chinese last name + foreign first name, and 2% were foreign last name + Chinese first name. The 28% rare Chinese names reduces the error in our assumption to 1.3%.

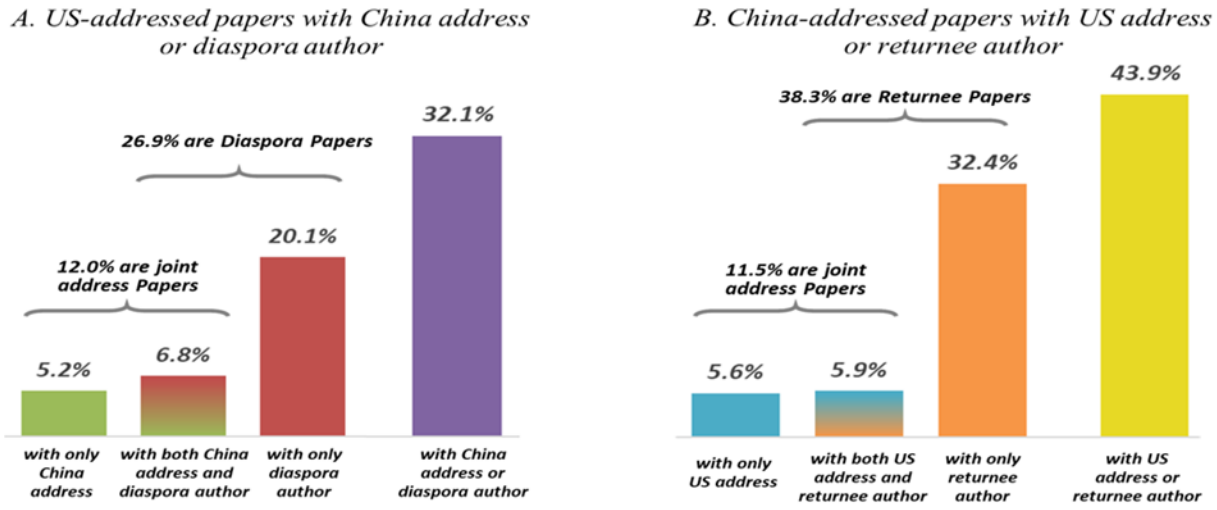
[9] Scopus Author ID are unique identifiers assigned by Scopus to differentiate authors with similar names. Aman (2020) and Conchi & Michels (2014) report that Scopus Author ID is a powerful tool in author name ambiguation.

with US-addressed and China addressed co-authors. Without gainsaying the importance of jointly addressed collaborations, the main way China-born researchers contribute to US science is by working in the US. Focusing on papers with no over-lapping groups makes the point vividly: the share of diaspora papers with *no* China addresses is nearly four times (3.9 = 20.1/5.2) the share of jointly addressed papers with *no* diaspora author.

Panel B shows the importance of returnee authors –authors with a US-addressed paper who later publish at a China address. The share of China-addressed papers with a returnee author far exceeds the share of China-addressed papers with US-addressed co-authors. Eliminating overlapping groups, the share of returnee papers with no US-addressed collaborators exceeds US-China joint addressed papers with no returnee author by nearly sixfold (5.8 = 32.4/5.6).

Taken together, the evidence in the two panels of Figure 1 demonstrates that the main channel by which China-born scientists collaborated with US-experienced scientists was through the **cross-country mobility of China-born researchers** to the US (diaspora authors) and their **return mobility** to China (returnee authors). In 2018, diaspora researchers were present on 26.9% of US addressed papers while returnee authors were present on 38.3% of China addressed papers, both far larger than the 11.5%-12% of joint addressed papers inclusive of diaspora and returnees.

Figure 1. The Diaspora Share of US papers and Returnee Share of China papers in 2018



Note: Appendix Table A3 gives detailed numbers.

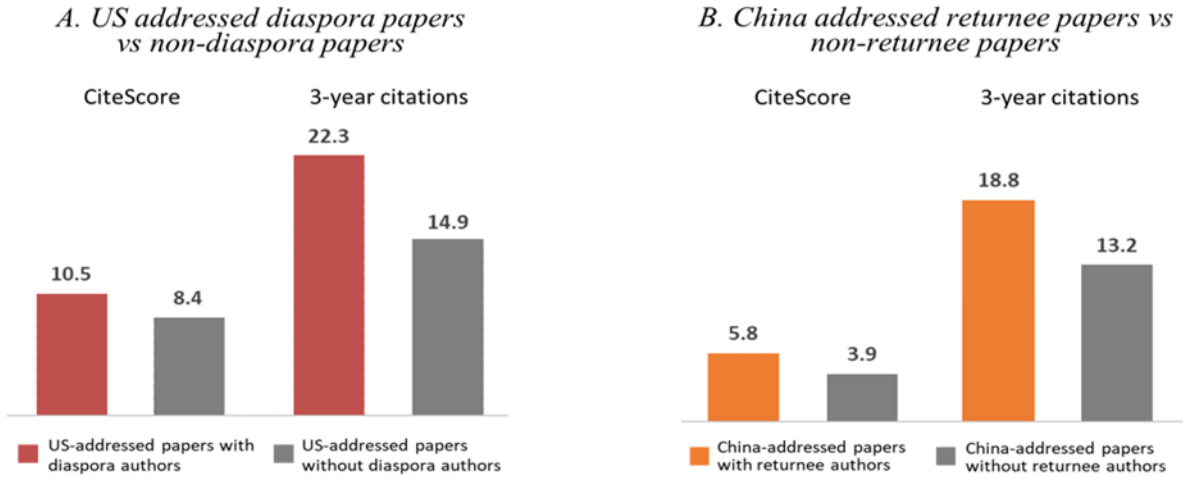
The scientific quality/impact of diaspora and returnee research

Going beyond numbers, papers with diaspora or returnee authors evince higher quality/impact than other US-addressed or China-addressed papers, respectively, as reflected in the number of citations received by the journal which published the paper before its publication; and the number of citations the paper itself receives after publication (Lariviere & Sugimoto, 2019). For the impact

of the journal of publication, we use Scopus's *CiteScore*[10] – For citations, we use 3-year forward citations received by 2018 papers from 2018-2021 Scopus publications. Given the life cycle of citations (Hajra & Sen, 2005; Wang et al, 2013; Stegehuis et al, 2015), 3 years is a short period for assessing citations, but in our data 3-year citations are highly correlated with citations over a longer period,[11]and thus a good indicator of longer-term citations. CiteScore and 3-year forward citations are correlated at 0.44 for all US-addressed papers and at 0.51 for all China-addressed papers in our data. This suggests that while the two variables are sufficiently correlated to reflect similar phenomenon, they can be viewed as a single "quality/impact" factor only with due allowance for measurement error.

Figure 2A shows that 2018 *diaspora papers scored above non-diaspora papers* in both CiteScores and citations, averaging 2.1 (25%) CiteScore points more and 7.4 (50%) 3-year citations more per paper than non-diaspora papers. Similarly, Figure 2B shows that *returnee papers scored above non-returnee papers,* averaging 1.9 (49%) CiteScore points more and 5.6 (42%) 3-year citations more than non-returnee papers.

Figure 2. CiteScore of journal of publication and citations from 2018-2021 publications for US and China-addressed papers, by diaspora and returnee status, 2018



Note: Appendix Table B1 & B2 give detailed numbers.

As factors beyond diaspora/returnee authorship influence CiteScore and citations, we estimated regression models that included variables that might account for part or all of the Figure 2 differences: the number of authors (more authors increases citations[12]), dummy variables for

[10] CiteScore is stable year to year: the correlation of CiteScore between 2020 and 2019 is 0.97; between 2020 and 2018 is 0.94; between 2020 and 2017 is 0.91.
[11] The correlation of 3-year citations with 7-year citations to the 2015 papers in our data is 0.98. (0.99 for 2015 USO papers, 0.83 for 2015 CO papers and 0.94 for 2015 ROW papers)
[12] Wuchty et al (2007)

fields, (fields differ in numbers of papers and/or in citing conventions[13]); and dummy variables for the different types of papers differentiated by addresses (due to national homophily of citations)[14].

Table 1 links CiteScore and citations to whether a paper is diaspora or is returnee and to the other variables connected to CiteScore and citations. To isolate the impact of diaspora researchers from any other China connection, columns 1 and 2 analyze US-addressed papers with no China address (USO and US-ROW). Similarly, to isolate the impact of returnee researchers in China from any other US connection, columns 3 and 4 analyze China addressed papers with no US address (CO and C-ROW). Columns 5 and 6 shows the estimated diaspora and returnee effects on papers with both US and China addresses.[15]

The column 1 and 2 regressions show that diaspora papers obtain significantly higher CiteScores and 3-year forward citations than other US-addressed papers in the presence of the measured attributes of the papers. The estimated differentials are noticeably smaller than the mean differences in Figure 2: an advantage of 1.1 in CiteScore compared to 2.1 in Figure 2 and an advantage of 5.3 in citations compared to 7.4 in Figure 2. The column 3 and 4 regressions show similarly that returnee papers obtain significantly higher CiteSores and 3-year forward citations than other China-addressed papers in the presence of the measured attributes, though with magnitudes roughly half the mean differences in Figure 2: an advantage of 0.99 in CiteScore compared to 1.9 in Figure 2 and an advantage of 2.8 in citations compared to 5.6 in Figure 2.[16]

By focusing on US and China collaborative papers, columns 5 and 6 allow us to assess the contribution of diaspora and returnee authors on papers in which both types of researchers appear. The estimated impacts of diaspora and returnee authors in these calculations exceed the estimated impacts in columns 1-4, which suggest that collaborative papers benefit from the presence of China-born authors at both addresses. While the estimated interaction of diaspora and returnee authors is insignificant, the sum of having a diaspora author and a returnee author adds 2.21 (1.55+1.35 -0.69) points to CiteScore and 14 (6.82+5.53+1.65) points to citations, so that papers with **both** diaspora and returnee authors top all papers. This suggests that diaspora and returnee researchers are largely complementary rather than substitute inputs in the underlying research.

[13] Schubert & Braun (1996) and Marx & Bornmann (2015)
[14] Schubert & Glänzel (2006) and Didegah & Thelwall (2013)
[15] The number of observations is smaller for CiteScores than for citations comes from papers that were published in newly established journals for which Scopus did not yet a CiteScore value. Regressions in which we pool all US-addressed and all China-addressed papers together yield similar results (Appendix Tables B7 and B8).
[16] The smaller diaspora advantage in the regressions for CiteScore and citations is due primarily to including the research field dummy variables in the regressions (Appendix Table B3). The smaller returnee advantage in the CiteScore regression is due to the inclusion of collaborative paper dummy variables while the smaller returnee advantage in the citation regression is due to inclusion of field dummies, author number, and international collaborative dummies roughly equally (Appendix Table B3).

As robustness checks on our findings, we estimated the impact of diaspora and returnee researchers on CiteScores and citations with alternative statistical models, given in Appendix Tables B4-8. In one set of regressions, we replaced the dummy variables for diaspora or returnee authors with the number of diaspora and returnee authors on a paper and obtained positive coefficients on the numbers of diaspora and returnees with magnitudes consistent with the Table 1 estimates (Appendix Tables B4). Given that the distributions of CiteScore and citations are upper tail skewed, we estimated regressions with Ln of CiteScore and Ln of 3-year citations as dependent variables, dropping the small number of papers with 0 citations, and confirm the statistically significant advantages of diaspora/returnee papers (Appendix Tables B5 & Appendix Tables B6).

Table 1. Regression Estimates of Effect of Diaspora and Returnee Papers on CiteScore and Citations of 2018 US and China addressed Papers

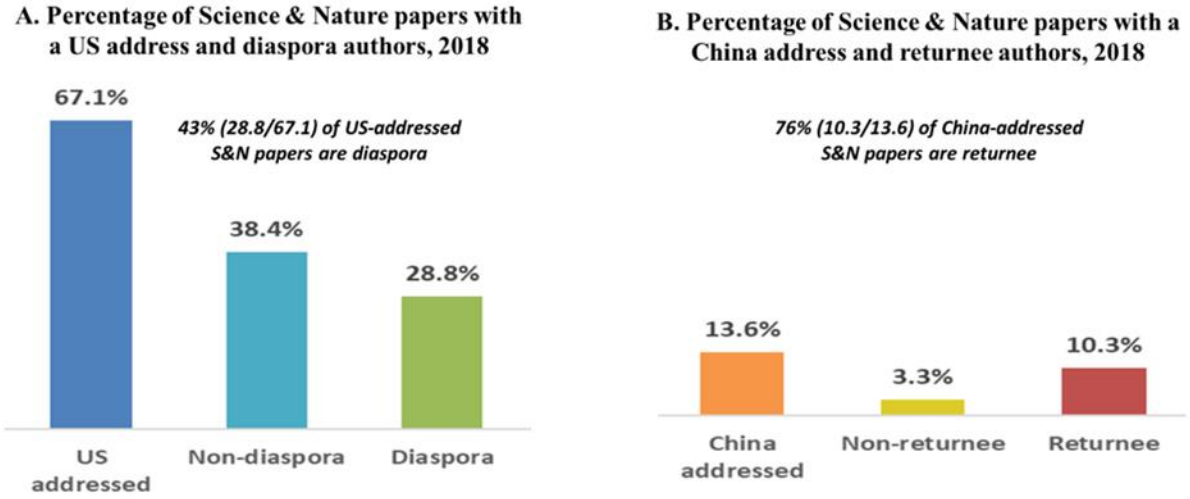
	US addressed papers with no China address (USO & US-ROW)		**China addressed papers with no US address (CO & C-ROW)**		**US and China collaborative papers (USC & USC-ROW)**	
	CiteScore	*Citations*	*CiteScore*	*Citations*	*CiteScore*	*Citations*
Diaspora paper dummy	1.11*** (0.226)	5.3*** (1.073)	-	-	1.55*** (0.263)	6.82*** (1.545)
Returnee paper dummy	-	-	0.99*** (0.099)	2.79*** (0.818)	1.35*** (0.257)	5.53*** (1.514)
Diaspora and Returnee dummy	-	-	-	-	-0.69 (0.513)	1.65 (3.016)
#Authors	0.04*** (0.004)	0.21*** (0.021)	0.17*** (0.012)	1.1*** (0.1)	0.05*** (0.007)	0.29*** (0.041)
US-ROW	0.2 (0.221)	2.91*** (1.048)	-	-	-	-
US-C-ROW	-	-	-	-	2.42*** (0.267)	11.9*** (1.566)
C-ROW	-	-	0.6*** (0.097)	3.7*** (0.802)	-	-
CO	-	-	--	--	-	-
Field dummy (21)	Yes	Yes	Yes	Yes	Yes	Yes
Mean of dependent variable	9.0	16.4	4.1	14.6	8.8	20.6
#Obs	6,345	6,533	3,801	4,000	3,782	3,908
Adjusted R2	0.1044	0.0385	0.1956	0.0698	0.1145	0.0458

Note: 95% confidence interval, * p<0.1; ** p<0.05; *** p<0.01. Standard errors are in brackets. Observations without valid address information, author information, or CiteScore value are dropped.

Finally, we examined the quality/impact of diaspora and returnee papers using their share of papers published in *Science* and *Nature* in 2018, taken together as S&N for ease of presentation. If the diaspora and returnee effects extend from the average papers to top papers, diaspora or returnee papers should be disproportionately represented in those (and other) leading journals. Figure 3 records the share of papers that were US-addressed papers with and without diaspora authors and the share of papers that were China-addressed with and without returnee authors. Panel A shows that 67% of S&N articles had at least one US address, of which 43% were diaspora

papers. This is 1.6 times of the diaspora share of all US-addressed papers reported in Figure 1. Panel B shows that 13.6% of S&N articles had at least one China address, of which 76% had a returnee. This is 2.0 times of the returnee proportion of all China-addressed papers in Figure 1.

Figure 3. Share of diaspora/returnee papers in Science and Nature 2018



Note: Calculated from all articles in Science and Nature in 2018, as described in Appendix Table B10. Diaspora papers averaged 159 3-year citations compared to 93 citations to non-diaspora papers. Returnee papers averaged 163 citations compared to 161.2 citations for non-returnee papers. These statistics suggest that having a China-addressed author boosted citations of S&N papers.

Taking citations, CiteScores, and presence on S&N papers as measures of quality/impact of research, these calculations indicate that the Figure 1 measures of shares of papers **understates** the contribution of diaspora researchers to US-addressed publications and of returnee researchers to China-addressed publications. Adjusting numbers of papers for quality/impact reflected in citations or CiteScores increases the 27% diaspora share of US papers to 38% in terms of citations and to 31% in terms of CiteScores, and increases the 38% returnee share of China papers to 52% in terms of citations and to 48% in terms of CiteScores.[17] Given concerns in China about the quality of research (Xie et al, 2014; Wagner et al, 2020; Brainard & Normile, 2022), the positive impact of returnees on CiteScores and citations suggests that they help address the quality issue.

[17] Adjustments are based on coefficients estimated on the relation between diaspora and returnee papers on CiteScore and Citations given in Appendix Table B7 for the exact samples used to generate Figure 1.

2. Diaspora and Returnee Authors in US-China Collaborations and Citations

This section shows that diaspora and returnee researchers further contributed to the US-China research connection by being central nodes in the collaboration and citation networks.

US-China collaborations

To the extent that being China-born gives diaspora authors both knowledge and interest to collaborate with China-based researchers and that having a US research background gives returnees knowledge and interest to collaborate with US-based researchers, we expect diaspora and returnee authors to be predisposed toward working on US-China collaborative papers compared to non-diaspora and non-returnee researchers. We test this hypothesis in two ways.

First, using the Figure 1 paper data, we compare the distribution of authors by diaspora and returnee status on 2018 US-China collaborative papers with the distribution that would arise if we *randomly selected* US-addressed authors from a weighted pool of all US-addressed authors and *randomly selected* China-addressed authors from a weighted pool of all China-addressed authors. We weighted each authors' representation in the pool on the basis of the number of papers they published in 2018 fractionated by the number of co-authors on each paper. Weighting by numbers of papers gave authors with more publications a higher chance of being on a joint collaboration, Fractionating authorship by the number of co-authors gave authors with more co-authors a smaller chance of being on a collaborative paper. As the table note reports, the results hold without any weighting because the distribution of papers and co-authors are similar for the various groups.

Table 2 compares the actual distribution of authors among collaborative papers (column 1) with the expected percentages if US-addressed authors were randomly selected from the weighted pool of US-addressed authors and if China-addressed authors were randomly selected from the weighted pool of China-addressed authors (column 2). It divides papers by authorship into four groups: those with at least one diaspora author; those with at least one returnee author, those with at least one diaspora **and** at least one returnee author; and those with at least one diaspora author **or** at least one returnee author. Column 3 shows that in all cases the actual percentage substantially exceeds the percentages from random selection.

The statistic that arguably best captures the significance of diaspora and returnee authors to US-China collaborations is the 78.5% of collaborative papers that have at least one diaspora or returnee. This falls short of the 100% that would indicate that a diaspora or returnee author is *necessary* for a US-China collaboration but is sufficiently high to potentially justify a term like *nearly necessary*.[18]

[18] Examining the 21.5% of US-China collaborations with neither a diaspora or returnee author, we found that 14% of the papers in our sample had a US-addressed author with a Chinese last name but non-Chinese first name, which would likely predispose them toward a US-China collaboration. We suspect that authors on some papers had links to the other country via schooling, being students or colleagues of returnee or diaspora researchers on which the bibliometric data is silent.

Table 2. Actual vs Predicted Distribution of US-China Collaborative Papers in 2018, by Presence of Diaspora (D) or Returnee (R) authors

D or R group	US-China collaborative papers		
	1. Actual %	2. Predicted % by random selection	3. Ratio of actual to predicted
With at least one D author	56.9%	28.0%	2.0
With at least one R author	51.1%	29.5%	1.7
With at least one D **and** at least one R authors	29.4%	6.9%	4.3
With at least one D **or** at least one R author	**78.5%**	50.6%	1.6

Note: Pool of potential authors based on numbers of 2018 papers of each author fractionated by the number of co-authors on each paper. The D share of fractionated US-addressed authors is *12%,* so the probability of at least one diaspora author on a US-China collaboration with ***n*** US-addressed authors is $1-(1-0.12)^{n}$. The R share of fractionated China-addressed authors is *8.5%,* so the probability of at least one returnee authors on a US-China collaboration with ***nc*** China-addressed authors is $1-(1-0.085)^{nc}$. Because the number of papers and co-authors on papers is similar among the groups the results hold without the weighting scheme. (Appendix Table C1, C2, and C3).

For our second test of the hypothesis that diaspora and returnee authors have a greater proclivity for writing US-China collaborations than non-diaspora and non-returnee autjhors, we randomly sampled 8,000 2018 authors and computed the share of **all** of their 2018 papers that were US-China collaborations.[19]

Table 3 shows that diaspora and returnee authors averaged higher proportions of US-China collaborations in their 2018 papers than their non-diaspora and non-returnee comparators. Column 1 displays this in terms of the mean percentage of authors' papers that are US-China collaborations with each author treated as an observation. Because the distribution of papers is far from normal for authors with small numbers of publications (authors with just one paper have either a 0% or 100% percentage collaboration), Column 2 records the mean of the authors' collaborative percentage weighted by the number of papers they published in 2018. This is equivalent to taking the ratio of the total number of US-China collaborative papers divided by the total number of papers by those authors. Column 3 uses the proportion of authors with *at least one* collaborative US-China paper as a dichotomous measure that identifies authors who participated in any 2018 US-China collaboration (which more resembles the Table 2 statistics on papers). All three of measures confirm the finding that diaspora and returnee authors were far more likely to work on US-China joint collaborations with researchers addressed in the other country than non-diaspora or non-returnee authors.

[19] The data set for this analysis consists of 53,197 papers published in 2018 by the 8,000 authors, as described in Appendix C4.

Table 3. Measures of 2018 Papers That Were US-China Collaborations, by Diaspora and Returnee Status of Authors

Type of Authors	1. Mean of Pct of 2018 papers that were US-China collaboration, by author	2. Ratio of Sum of US-China collaborations to all 2018 papers*	3. %Authors with at least one US-China collaborative papers in 2018
US Addressed Authors			
Diaspora	26.0%	34.3%	46.8%
Non-diaspora	6.7%	9.4%	18.4%
Ratio (D/ND)	*3.9*	*3.7*	2.5
China Addressed Authors			
Returnee	22.9%	20.4%	65.4%
Non-returnee	8.8%	8.3%	24.2%
Ratio (R/NR)	*2.6*	*2.5*	2.7

Note: See Appendix Table C4 for details about the sampled authors. *This column is equivalently the paper-weighted average of authors ratios.

Finally, we examine the extent to which authors wrote US-China collaborative papers before 2018. The most visible group of such researchers are researchers holding appointments in both countries at the same time, whose papers are all presumptively US-China collaborations. Based on our data in 2018 there were 12,919 authors with dual affiliations in 2018. Despite the dual affiliated authors being a small share of all Chinese-addressed and US-addressed authors on collaborative papers, we estimate that they are present on 21.5% of collaborative papers in 2018. Confirming the importance of China-born researchers in the research connection, 90.8% of dual addressed authors had both Chinese first and last names. (See Appendix Table D1 for details).

Going beyond authors with a dual affiliation in 2018, we next show that diaspora authors in 2018 were more connected to research in China than non-diaspora authors prior to 2018 and thus can be viewed as providing a relatively long-term link between US and China research. Table 4A shows huge differences between diaspora and non-diaspora researchers in the percentages of publications where they have a China address or where they have a US address but have China addressed co-authors. Table 4B shows a similar pattern for returnee authors compared to non-returnee authors. In this case, column 1 shows that returnees had a US address on 11.4% of all their pre-2018 publications, while by definition non-returnees had a US address on 0% of all their pre-2018 publications. 2018 returnee authors were 2.1 times more likely to co-author with a US addressed researcher in pre-2018 publications than 2018 non-returnee authors. In total, returnees were 4.5 times more likely to have a US connection than non-returnees.[20]

[20]Dual addressed authors with at least one pre-2018 publication show the strongest link between China and the US over time, with 71% having a dual address on at least one pre-2018 publication, 89% having at least one solo-China pre-2018 address, and 61% having at least one solo-US pre-2018 address and 79% having at least one paper with their address solely in one country and a coauthor from the other country (See Appendix Table D2 for details).

Table 4: Percent of Pre-2018 Publications with Connection to Other Country by 2018 US-Addressed and China-Addressed Authors, by Diaspora and Returnee Status

A. % of pre-2018 publications by 2018 US-addressed authors where *the author had*			
	China address for themself	US address for themself but China addressed coauthors	Any China connection
Diaspora	12.4%	17.3%	29.7%
Non-diaspora	0.3%	3.7%	4.0%
B. % of pre-2018 publications by 2018 China-addressed authors where *the author had*			
	US address for themself	China address for themselves but US addressed coauthors	Any US connection
Returnee	11.4%	10.5%	21.9%
Non-returnee	0.0%	4.9%	4.9%

Note: Any China (US) connection is sum of percent of papers where the author had the other country's address and other country addressed co-authors. For simplicity, we count US-China dual address as China address in Panel A and count US-China dual address as US address in Panel B. See Appendix Table D3 for details.

Diaspora and returnee researchers in US-China citations

Do the close ties between diaspora authors and China-addressed authors and between returnee authors and US-addressed authors in collaborations extend to the network of citations?

To determine whether US-addressed papers with diaspora authors cite papers differently than papers without diaspora authors and, commensurately, whether China-addressed papers with returnee authors cite papers differently than those without returnee authors, we compared the citation behavior of authors who published papers in 2016-18 toward papers published in 2015 (so the citations are three-year forward citations to 2015 publications). If diaspora researchers are closer than non-diaspora researchers to Chinese-based research, 2016-18 diaspora papers should cite China-addressed publications compared to non-China-addressed Rest of World (ROW) publications more than do 2016-2018 non-diaspora papers. To identify the diaspora/non-diaspora difference in citing behavior independent of any other addressed-based connection between authors, we limit analysis to US-addressed papers with US Only (USO) addresses and to China-addressed papers with China Only (CO) addresses. This removes US-China, China-ROW and US-ROW joint addressed papers from our analysis. Since the analysis compares the citing behavior of diaspora papers between China-addressed and ROW publications to the citing behavior of non-diaspora papers between China-addressed and ROW publications, it is a double-difference comparison, with the first difference being the ratio of citations from diaspora papers to CO papers compared to ROW papers and the second difference contrasting the diaspora ratio to the analogous ratio of CO to ROW of citation ratios made by non-diaspora papers.

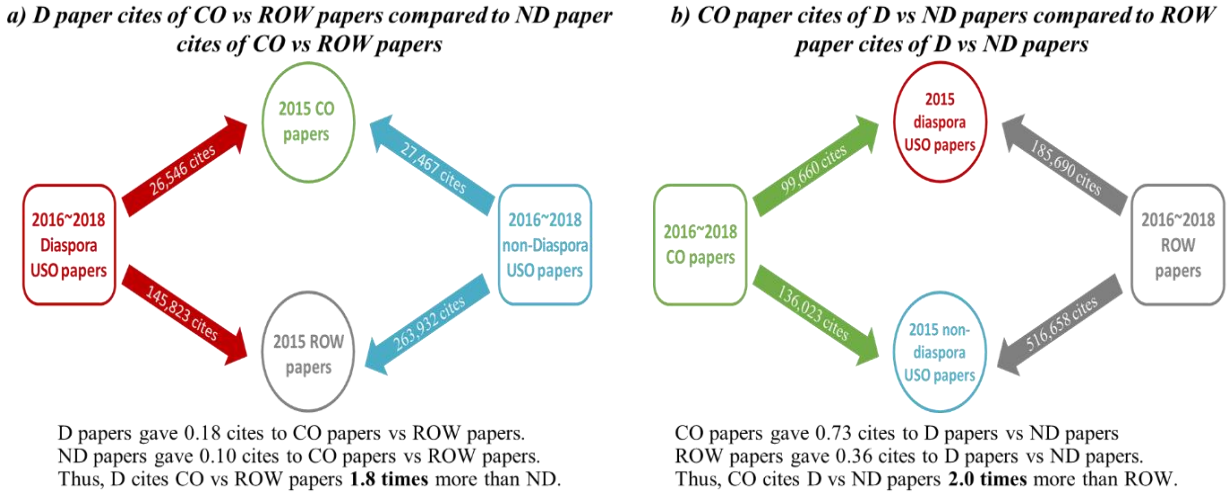
All of empirical analyses are based on the citation data set described in Appendix D4.

Figure 4a shows our calculation for the citing preference of diaspora and non-diaspora papers. In our data set, 2016-18 *diaspora USO* papers gave 26,546 citations to 2015 CO papers compared to 145,823 citations to the more numerous ROW papers for a first difference citation ratio of 0.18. By comparison *non-diaspora USO* papers gave 27,647 citations to CO papers compared to 263,932 citations to ROW papers, for a ratio of 0.10. The "second difference" ratio of the 0.18 diaspora paper preference to the 0.10 non-diaspora paper preference shows that diaspora papers have a 1.80 preference for citing CO papers compared to ROW papers relative to non-diaspora papers.

Using the same methodology, Figure 4b shows that 2016-18 CO papers cite diaspora papers compared to non-diaspora papers by a ratio of 0.73 compared to an ROW ratio of citations between diaspora and non-diaspora papers of 0.36. This gives a 2.0 times preference of CO to ROW citations for diaspora vs non-diaspora papers.

In short, diaspora USO papers have a preference for CO papers in their citing behavior and CO papers have a preference for diaspora USO papers.

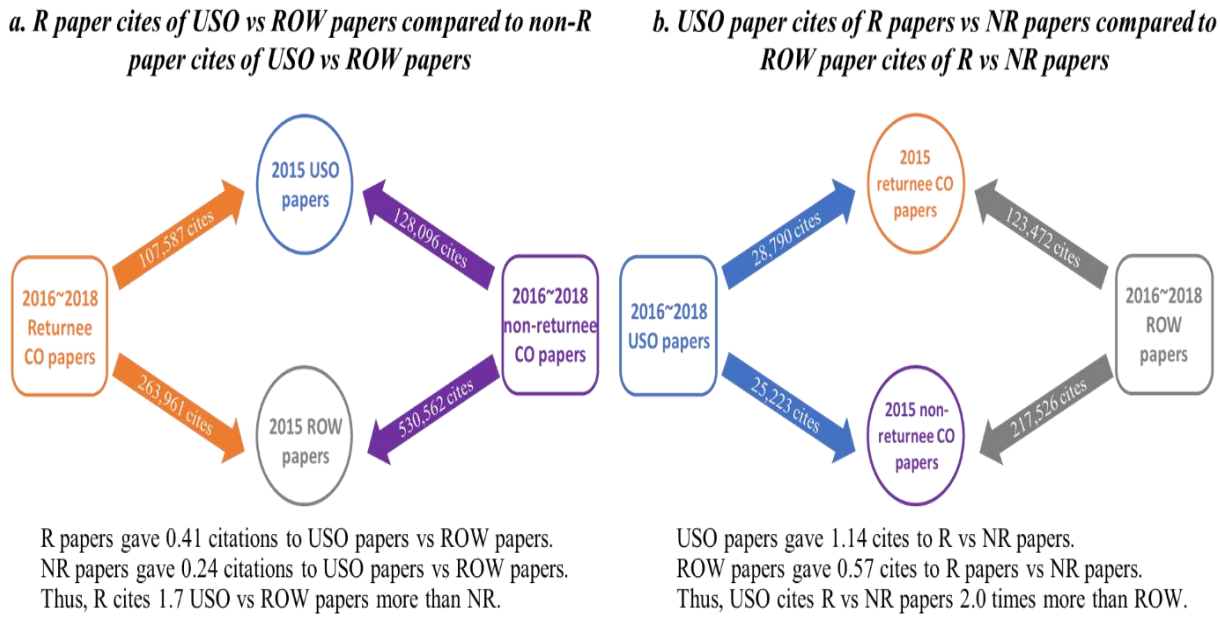
Figure 4. Citations from 2016-18 Papers to 2015 Papers, authors and addresses on papers



Source: see Appendix Table D4 for details

Figure 5 uses the same double difference analysis to contrast the citing behavior of returnee CO papers compared to non-returnee CO papers. In Figure 5a, the first difference is in the citing behavior of returnee papers between USO and ROW papers. The second difference contrasts the returnee citing behavior to the non-returnee citing behavior between USO and ROW papers. The final result is a 1.7 differential preference of returnee to non-returnee citations toward USO papers vs ROW papers. In Figure 5b, the first difference is in the citations that USO papers give to CO returnee papers relative to CO non-returnee papers. The second difference is between the USO citation preference to the preference of ROW papers to returnee papers vs non-returnee papers. The differential ratio preference is 2.0. In short, returnee papers disproportionately cite USO papers and are disproportionately cited by USO papers.

Figure 5. Three-year Citations from 2016-18 Returnee and USO papers to 2015 Papers: and Returnee and USO Papers Give More citations to each other



Source: see Appendix Table D4 for details

Taken together, the citing preferences between diaspora USO papers and CO papers and between returnee CO papers and USO papers shows that diaspora and returnee authors were key nodes in the flow of citations between the US and China. Contrary to Kipling's Ballad of East and West (1886) that "East is East and West is West and never the twain shall meet" diaspora and returnee researchers bridged the differences between the US and China in 2018 to link the two countries in research papers, collaborations, and citations into what could be described as effectively a single research community.

3. Breaking the Twain of the US-China Research Connection?

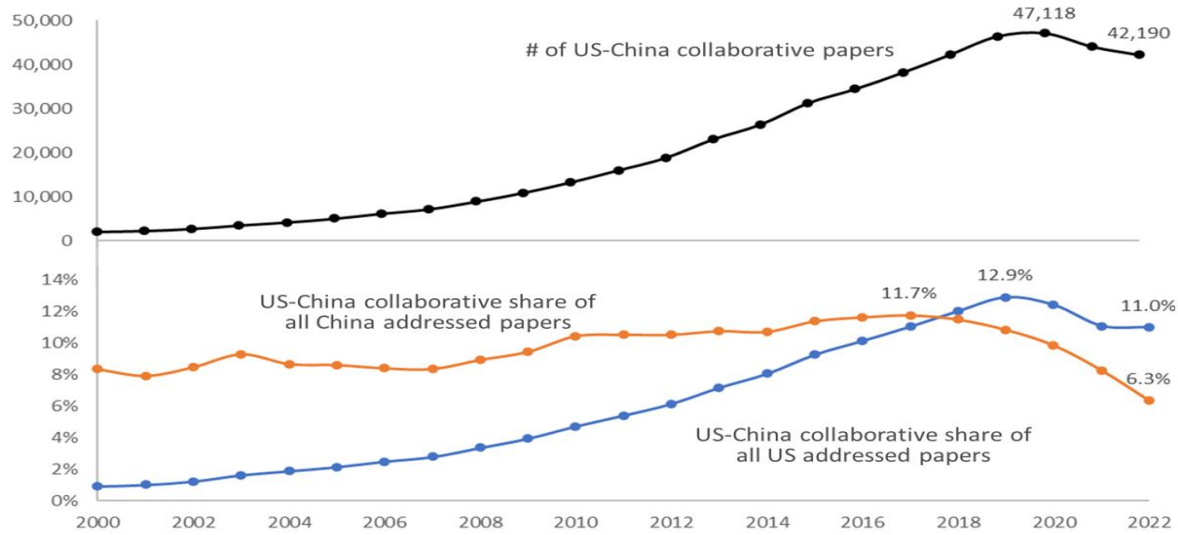
Success notwithstanding, the US-China research connection came under substantial strain in the late 2010s to early 2020s as the US and Chinese governments came to view each other more as geo-political adversaries than research partners and as, politics aside, the COVID-19 pandemic reduced the flow of students and researchers between the countries. This section examines how the strain showed up in measures of scientific ties and what those changes portend for the future.

Fraying of ties: Declines in US-China collaborations

Figure 5 gives the first indicator of a fraying of ties post-2018. The upper panel shows the upward trend in US-China collaborative papers slowing in the late 2010s and then falling by 10.5% from 2020 to 2022. Given lags between research activity and publication, the timing of the decline

is consistent with the Trump Administration's 2017-2021 "*China Initiative*"[21] and also with the onset of the COVID-19 pandemic. While the Biden Administration ended the China Initiative, it maintained national security concerns about collaborations with China and introduced the 2022 CHIPS and Science Act in part to "counter China". American government wariness of China combined with Xi Jinping's 2022 stress on "*self-reliance and strength in science and technology*" made US-China scientific collaborations more difficult than in the past.[22] The lower panel shows that the faster growth of papers in China than in the US translated into a much larger drop in the US-China collaborative share of China papers – 5.4 percentage points from 2017 to 2022 – than of US papers – 1.9 percentage points from 2019 to 2022. From this perspective, China's research separated more from US research than did US research from China's research.

Figure 5. Papers with US and China addresses and their shares of all US and China papers



Source: Scopus database, 2019-2022 data collected at March 2023.

Even with the two governments "leaning against" collaborations, however, it will take much greater drops in joint work to topple the US and China from leading the world in collaborative papers. In 2022 US-China joint papers exceeded by 74% the number of joint papers in the world's 2nd biggest collaboration.[23] Even in AI, a major area of governmental concern to national security and economic interest, the US and China were each other's top international partner in research, far outpacing each's 2nd place international collaborator (see Maslei, et al, 2023, Figure 1.1.6 & 1.1.7). Barring a huge deterioration in US-China relations that would drastically disrupt and re-orient international collaborative research in both countries, the near-term future is that China and the US will remain close international collaborators. Per gravity models of collaborative work in

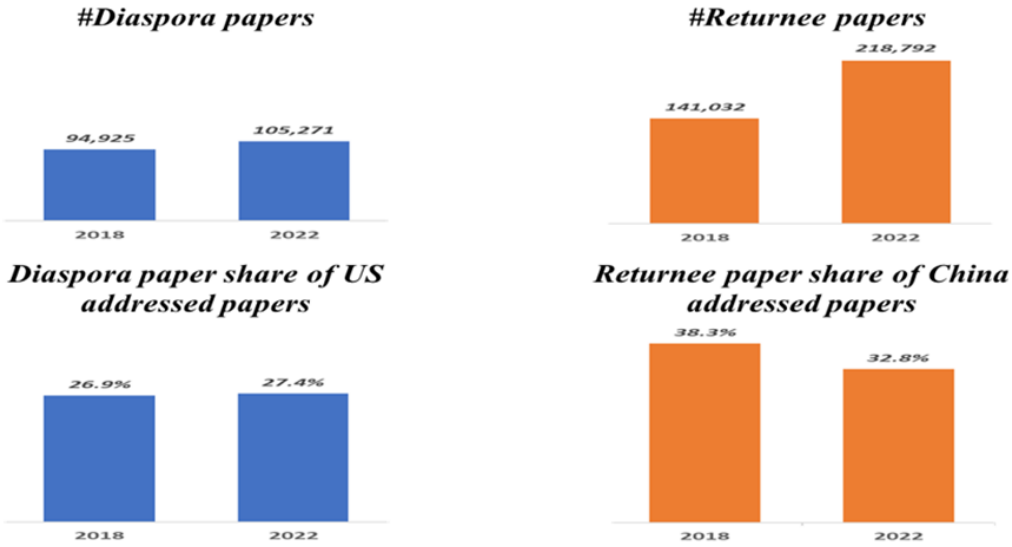
[21] https://www.uscc.gov/research/timeline-executive-actions-china-2017-2021
[22] See https://www.whitehouse.gov/briefing-room/statements-releases/2022/08/09/fact-sheet-chips-and-science-act-will-lower-costs-create-jobs-strengthen-supply-chains-and-counter-china/. Xi Jinping's statement is from http://english.www.gov.cn/news/topnews/202210/25/content_WS6357df20c6d0a757729e1bfc.html.
[23] 42,190 S&E English language articles had US and China addresses compared to 24,230 that had US and UK addresses.

which the magnitude of country activity increases collaborations[24], China's continued increase in global scientific activity will likely raise the number of researchers seeking US collaborations, with different effects on the collaboration shares of the two country's papers depending on relative growth rates of papers and changes in the potential for working with other countries.

Diaspora and Returnee Papers

Figure 6 shows that the number of diaspora and returnee papers that our study has identified as the main pathway for China-born researchers to connect with US-based research increased in the period. Between 2018 and 2022, the number of diaspora papers increased moderately (11%) while the number of returnee papers increased greatly (55%).[25] The moderate increase in diaspora papers raised its share of US papers by 0.5 percentage points while the huge increase in China papers dwarfed the increase in returnee papers so the returnee share fell by 5.5 percentage points.

Figure 6. Numbers and Shares of Diaspora and Returnee Papers, 2018 and 2022.



Source: Scopus database, 2022 data collected at March 2023.

International Scholars and Students

Given that many Chinese students and researchers come to the US while few US students and researchers go to China, we examine in Table 5 the flow of Chinese scholars between China and the US upon which the research connection rests. Panel A records the number of students and scholars enrolled in US educational institutions -- a "stock" measure that depends on

[24] Gravity models link collaborations to a multiplicand of the size of each country's scientific activity (relative to the distance between them). (Zhang & Guo 2017; Avdeev, 2021. Micro-based matching models (https://en.wikipedia.org/wiki/Search_and_matching_theory_(economics) give a natural bound on collaborative papers by the size of the smaller group.

[25] Consistent with this, OECD (2023) estimates that the "*net flow of scientific authors*" based changes in their country address turned from highly positive in 2015 to negative in 2021 for the US while increasing for China in 2021 (Figure 2.9). The share of Chinese nationals with new US PhDs intending to stay in the US also fell in 2021 (NSF, October 2022. Table 2-8).

admissions over several years and stay/leave decisions by admitted students. Prior to the pandemic, Chinese citizens made up about 1/3rd of US international students and scholars, by far the largest country group. The 2020/21 pandemic reduced the number of international students and scholars from all countries by roughly the same proportion. In 2021/22, however, while the total number of international students and scholars began to recover the number from China kept falling. The most likely reason is that China maintained its "Zero COVID" policy, through December 2022, which made domestic and international travel difficult through 2021/22.

Panel B gives the number of F1 (student) and JI (researcher) visas issued by the US. These are more volatile "flow" measures, which fell massively in 2020, particularly for China. When the US issued more visas in 2021, China's share of F1 visas recovered to its 2018-19 level, but then fell in 2022, most likely due to the "*Zero-COVID*" policy,[26] with China losing its spot as number one country in F1 visas to India. With the Dec 2022 end of the Zero-COVID policy, however, the number and percentage of F1 visas to Chinese citizens jumped to pre-pandemic levels in Q1 of 2023 (see Appendix Table E1), making China number once again top in student visas to US. The number of J1 visas fell more sharply to just 2.3% of J1 visas in 2022, and recovered slowly in Q1 2023, possibly because the limited number of flights and high cost of tickets between the countries[27] discouraged short to medium term research visits (see Appendix Table E2).

Table 5. Chinese International students and scholars in the US and F1 & J1 visas 2018-2022

Panel A. Number of Chinese International Students and Scholars						
Academic year	**Total # of international students in the US**	**# of international students in the US from China**	**% from China**	**Total # of international scholars in the US**	**# of international scholars in the US from China**	**% from China**
2017/18	1,094,792	363,341	33.2	135,009	46,256	34.3
2018/19	1,095,299	369,548	33.7	136,563	47,964	35.1
2019/20	1,075,496	372,532	34.6	123,508	42,863	34.7
2020/21	914,095	317,299	34.7	85,538	26,254	30.7
2021/22	948,519	290,086	30.6	90,891	19,391	21.3
Panel B. Number of F1 (student) and J1(scholars and others) visas issued to Chinese citizens						
Calendar year	**Total # of F1 visas issued by the US**	**# F1 visas issued to Chinese citizens**	**% to China**	**Total # of J1 visas issued by the US**	**# J1 visas issued to Chinese citizens**	**% to China**
2018	359,859	97,683	27.1	345,546	39,109	11.3
2019	363,607	98,584	27.1	353,023	39,167	11.1
2020*	102,850	4,853	4.7	63,246	1,925	3.0
2021	391,041	99,431	25.4	166,390	4,676	2.8
2022	409,156	57,511	14.1	293,973	6,849	2.3

Source: Institute of International Education (2023) and Monthly Nonimmigrant Visa Issuance Statistics from U.S. Bureau of Consular Affairs, accessed at April 2023. * Note that the IIE data "*include students on a temporary non-immigrant visa, regardless of if the student was physically located in the United States.*" (IIE, 2023).

[26] The US consulate in Shanghai that normally issues many visas was closed in April 2022 when Shanghai was locked down.
[27] Reuters (2023) reports only 72 flights between the US and China in Jan 2023 compared to 2961 flights per month in 2019.

Conclusion

Identifying Chinese diaspora researchers in the US by their names and returnee researchers by their publication history, our study has found that diaspora and returnee researchers contributed to the quantity and quality/impact of papers in both countries, were part of most US-China collaborations, and were key nodes in the network of citations that connects research in the two countries. While US-China political discord and economic competition and the COVID-19 pandemic frayed the research connection by reducing US-China collaborations and mobility of researchers and students, diaspora and returnee researchers maintained their links with the other country, providing a relatively permanent channel for scientific communication and collaboration.

To the extent that the research communities in both countries – scientists, universities, firms and students – continue to find value in the US-China research connection, and that governments take account of the benefits that diaspora and returnee research has brought to both countries, the connection is likely to remain a major part of global science in the foreseeable future. Rational decision-making favors US and Chinese researchers working together (and with researchers in other countries) in areas of potential existential threat to humanity: global warming and zoonotic pandemics, and of finding ways to produce sustainable economic growth that reduces poverty and of ways to end conflicts without war, as well as producing scientific knowledge orthogonal to current practical concerns but that can help us address unanticipated future dangers to well-being. Arguments for de-coupling technologies, shortening supply chains, and protecting some knowledge for national security reasons, while potentially valid in an era of global rivalry, apply less to scientific research than to almost any other human activity.

Reference

1. Aman, V., 2020. Transfer of knowledge through international scientific mobility: introduction of a network-based bibliometric approach to study different knowledge types. Quantitative Science Studies, 1(2), pp.565-581.

2. Avdeev, S., 2021. International collaboration in higher education research: A gravity model approach. Scientometrics, 126(7), pp.5569-5588.

3. Brainard, J. and Normile, D., 2022. China rises to first place in one key metric of research impact. Science, 377(6608), pp.799-799.

4. Chinese Ministry of Public Security, 2020. The 2019 National Name Report (二〇一九年全国姓名报告). Available at https://www.mps.gov.cn/n2254314/n6409334/c6874817/content.html.

5. Didegah, F. and Thelwall, M., 2013. Which factors help authors produce the highest impact research? Collaboration, journal and document properties. Journal of informetrics, 7(4), pp.861-873.

6. Freeman, R.B. and Huang, W., 2015. Collaborating with people like me: Ethnic coauthorship within the United States. Journal of Labor Economics, 33(S1), pp.S289-S318.

7. Hajra, K.B. and Sen, P., 2005. Aging in citation networks. Physica A: Statistical Mechanics and its Applications, 346(1-2), pp.44-48.

8. Institute of International Education, 2023. Open Doors Reports 2023. Available at https://opendoorsdata.org/.

9. Lariviere, V. and Sugimoto, C.R., 2019. The journal impact factor: A brief history, critique, and discussion of adverse effects. Springer handbook of science and technology indicators, pp.3-24.

10. Lin, W.C. and Chang, C.W., 2022. The influence of Chinese scholars on global research. Scientific Reports, 12(1), p.18410.

11. Marx, W. and Bornmann, L., 2015. On the causes of subject-specific citation rates in Web of Science. Scientometrics, 102, pp.1823-1827.

12. Maslej, N., Fattorini, L., Brynjolfsson, E., Etchemendy, J., Ligett, K., Lyons, T., Manyika, J., Ngo, H., Niebles J.C., Parli, V., Shoham, Y., Wald, R., Clark, J. and Perrault, R., 2023. The AI Index 2023 Annual Report. AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, CA.

13. National Center for Science and Engineering Statistics, National Science Foundation, 2022. Doctorate Recipients from U.S. Universities: 2021. NSF 23-300. Alexandria, VA. Available at https://ncses.nsf.gov/pubs/nsf23300.

14. National Science Board, National Science Foundation, 2021. Publications Output: U.S. and International Comparisons. Science and Engineering Indicators 2022. NSB-2021-4. Alexandria, VA. Available at https://ncses.nsf.gov/pubs/nsb20214/.

15. OECD, 2023. Science, Technology and Innovation Outlook 2023: Enabling Transitions in Times of Disruption, figure 2.9. OECD Publishing, Paris. Available athttps://doi.org/10.1787/f3c247fc-en.

16. Reuters, 2023. Airlines face hurdles to cashing in on China re-opening. https://www.reuters.com/business/aerospace-defense/airlines-face-hurdles-cashing-china-re-opening-2023-01-17/.

17. Schubert, A. and Braun, T., 1996. Cross-field normalization of scientometric indicators. Scientometrics, 36(3), pp.311-324.

18. Schubert, A. and Glänzel, W., 2006. Cross-national preference in co-authorship, references and citations. Scientometrics, 69, pp.409-428.

19. Shi, D., Liu, W. and Wang, Y., 2023. Has China's Young Thousand Talents program been successful in recruiting and nurturing top-caliber scientists?. Science, 379(6627), pp.62-65.

20. Stegehuis, C., Litvak, N. and Waltman, L., 2015. Predicting the long-term citation impact of recent publications. Journal of informetrics, 9(3), pp.642-657.

21. Tang, L., Cao, C., Wang, Z. and Zhou, Z., 2021. Decoupling in science and education: A collateral damage beyond deteriorating US–China relations. Science and Public Policy, 48(5), pp.630-634.

22. Van Noorden, R., 2022. The number of researchers with dual US-China affiliations is falling. Nature, pp.235-236.

23. Wagner, C.S. and Cai, X., 2022. Changes in co-publication patterns among China, the European Union (28) and the United States of America, 2016-2021. arXiv preprint arXiv:2202.00453.

24. Wagner, C.S., Cai, X. and Mukherjee, S., 2020. China's scholarship shows atypical referencing patterns. Scientometrics, 124(3), pp.2457-2468.

25. Wang, D., Song, C. and Barabási, A.L., 2013. Quantifying long-term scientific impact. Science, 342(6154), pp.127-132.

26. Wuchty, S., Jones, B.F. and Uzzi, B., 2007. The increasing dominance of teams in production of knowledge. Science, 316(5827), pp.1036-1039.

27. Xie, Y., Zhang, C. and Lai, Q., 2014. China's rise as a major contributor to science and technology. Proceedings of the National Academy of Sciences, 111(26), pp.9437-9442.

28. Zhang, C. and Guo, J., 2017. China's international research collaboration: Evidence from a panel gravity model. Scientometrics, 113, pp.1129-1139.

Appendix

Appendix A: methodology and quantity of diaspora & returnee research

Method of distinguishing the diaspora papers

Appendix Figure A1 shows how we estimate the number of Chinese diaspora papers in 2018. We used the Scopus query string to download files that contain bibliographic data of papers from the Scopus online website[1]. The files include the paper's unique *EID*[2]; author information: last names and first name initials of authors, authors' addresses, *Author ID*[3]; and the number of citations received to the date of downloading. We determined the Chinese ethnicity of authors by whether their last names are common Chinese last names as reported in the Chinese Ministry of Public Security's list of Chinese last names[4]. By this count, 30.8% of papers with only US addresses had an ethnic Chinese author.

To differentiate mainland-born Chinese researchers from Chinese ethnic persons born outside the mainland on US-addressed papers, we divide authors with Chinese last names into those with Chinese first names indicative of being mainland born and those with non-Chinese first names. Because the online website files do not provide full *first names* of authors, we randomly sampled 2,000 papers of ethnic Chinese persons in each of our groups of US-addressed papers (USO, US-C, US-C-ROW, and US-ROW) and used the EIDs of sampled papers to retrieve author first names via the Scopus API portal. Appendix Table A1 gives the details of the sampled US addressed papers.

We use authors' first names to estimate the proportion of ethnic Chinese authors likely to have been born in mainland China based on whether their first names follow the grammar of the Hanyu Pinyin translation system used in mainland China[5]. Our scheme labels Jianguo Wang as mainland China born and John Wang as non-mainland China born. Because the structure of pinyin syllables used in mainland China differs from that in other Chinese language speaking areas, our program also differentiates mainland names from other Chinese language area names[6]. We identify authors with typical Chinese last and first names at a US address as a diaspora author in the US.[7]

---

[1] https://www.scopus.com

[2] See https://dev.elsevier.com/guides/ScopusSearchViews.html

[3] Author ID is the unique identifier Scopus assigns to differentiate authors with similar names, see: https://service.elsevier.com/app/answers/detail/a_id/11212/c/10546/supporthub/scopus/kw/AU-ID/.

[4] The 2019 National Name Report: https://www.mps.gov.cn/n2254314/n6409334/c6874817/content.html lists the most common Chinese last names. It covers the Chinese last names of 84.8% of the mainland population.

[5] The program that distinguishes Chinese first names is available at GitHub: https://github.com/qingnanxie/Chinese-first-name.

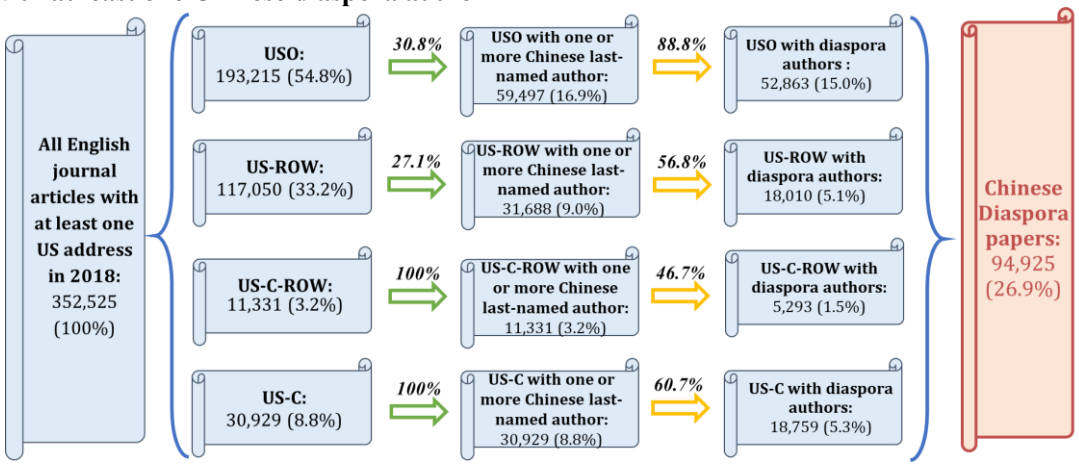
[6] For example, *Xie* is the mainland pinyin translation of 谢, which is translated as *Tes* in Hong Kong and *Hsieh* in Taiwan.

[7] Our methodology misses the likely small number of authors who changed their names into non-Chinese names, Chinese ethnicity authors born outside of China but given a Chinese name, and authors with rare Chinese names.

Our identification includes US-addressed authors who may also have a China or other country address.

Multiplying those sample-based proportions of the four types papers (listed on the yellow arrows in Appendix Figure A1) to the number of each type of papers respectively, we estimated the number of USO papers with diaspora authors, the number of US-ROW papers with diaspora authors, the number of US-China-ROW papers with diaspora authors, and the number of US-China papers with diaspora authors.

Appendix Figure A1. Estimated number of diaspora papers in 2018: #US addressed papers with at least one Chinese diaspora author



Note: Green arrows refer to calculations based on the query string results (population counts), yellow arrows refer to calculations based on sampling estimations.

Appendix Table A1. Samples for US-addressed papers

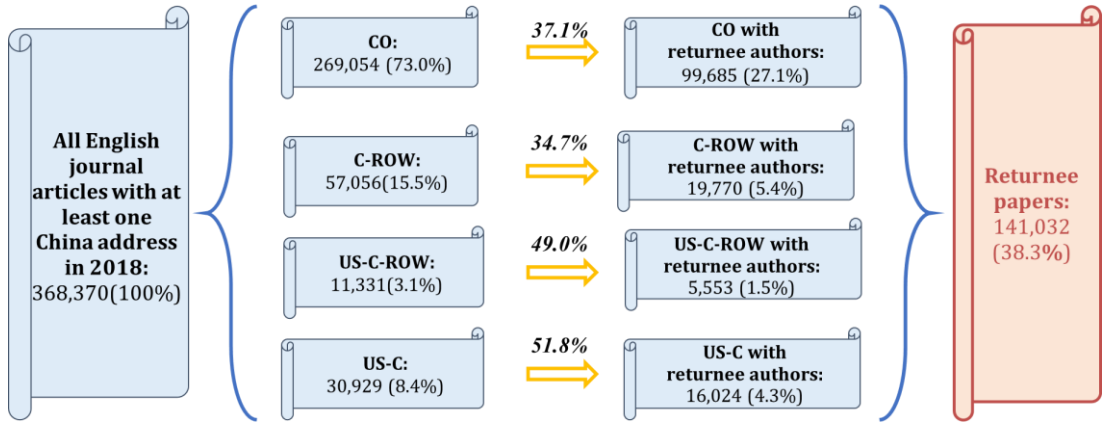
Data Sample	Purpose	Years Covered	Total number sampled
USO papers with Chinese last-named authors	・Estimate the number of USO papers with diaspora authors in 2015, 2016, 2017, 2018; ・Estimate the scientific impact of USO papers with diaspora authors (CiteScore & Citation for 2018 papers); ・Analyze the citing behavior of the 2016-2018 diaspora/returnee papers (3-year citation data of 2015 USO papers); ・Analyze diaspora and non-diaspora authors' co-authorship networks; ・Analyze the pre-2018 publications of diaspora and non-diaspora authors.	2015-2018	2,000 in each year for total of 8,000
US-ROW papers with Chinese last-named authors (No China address)	・Estimate the number of US-ROW papers with diaspora authors; ・Estimate the scientific impact of US-ROW papers with diaspora authors; ・Analyze diaspora and non-diaspora authors' co-authorship networks; ・Analyze the pre-2018 publications of diaspora and non-diaspora authors.	2018	2,000
US-C-ROW papers	・Estimate the quantity of US-C-ROW papers with diaspora authors; ・Estimate the quantity of US-C-ROW papers with returnee authors; ・Estimate the scientific impact of US-C-ROW papers with and without diaspora authors; ・Estimate the scientific impact of US-C-ROW papers with and without returnee authors; ・Analyze the role of diaspora authors and returnee-diaspora authors in the 2018 US-China collaboration; ・Analyze the co-authorship networks of diaspora, non-diaspora, returnee and non-returnee authors; ・Analyze the pre-2018 publications of diaspora, non-diaspora, returnee and non-returnee authors.	2018	2,000
US-C papers (No rest-of -world countries' address)	・Estimate the quantity of US-C papers with diaspora authors; ・Estimate the quantity of US-C papers with returnee authors; ・Estimate the scientific impact of US-C papers with and without diaspora authors; ・Estimate the scientific impact of US-C papers with and without returnee-diaspora authors; ・Analyze the role of diaspora and returnee-diaspora authors in the 2018 US-China collaboration; ・Analyze the co-authorship networks of diaspora, non-diaspora, returnee and non-returnee authors; ・Analyze the pre-2018 publications of diaspora, non-diaspora, returnee and non-returnee authors.	2018	2,000
USO papers	・Estimate the scientific impact of USO papers without diaspora authors	2018	2,000
US-ROW papers	・Estimate the scientific impact of US-ROW papers without diaspora authors	2018	2,000

Method of distinguishing the returnee papers

We estimated the number of China-addressed returnee papers in two steps, as shown in Appendix Figure 2. First, we searched English language journal articles for those with at least one China address and then divided those articles into: CO papers (China addresses only); C-ROW papers (China and rest-of-world countries addresses); US-C-ROW papers (US, China, and rest-of-world country addresses); and US-C papers (US and China addresses only). Appendix Table A2 gives the sampling details.

We sampled 2,000 of the papers in each group, giving a total sample of 8,000 papers, and took the Scopus Author ID associated with the papers to retrieve data on **all** Scopus-indexed publication of that author[8]. Our returnee authors are China-addressed authors who had an article with their name at a US address prior to their 2018 China-addressed article. The estimated proportion of China-addressed papers with a returnee author varied from 51.8% for US-C papers to 34.7% of C-ROW papers. Multiplying the number of 2018 papers of each group by the estimated proportions of papers with returnee authors and summing across the 4 groups gives our bottom-line estimate that 38.3% of 2018 China-addressed papers had at least one author with prior US research experience.

Appendix Figure A2. Estimated number of returnee papers in 2018: #China addressed papers with at least one returnee author



Note: Yellow arrows refer to calculations based on sampling estimations.

[8] We use all publications indexed in Scopus instead of just English journal S&E articles to capture as much as possible returnee authors.

Appendix Table A2. Samples for Additional China-addressed papers and ROW papers

Data Sample	Purpose	Years Covered	Total number sampled
CO papers	・Estimate the quantity of CO papers with and without returnee authors in 2015-2018; ・Estimate the scientific impact of CO papers with returnee authors (CiteScore & Citation for 2018 papers); ・Analyze the citing behavior of the 2016-2018 diaspora/returnee papers (3-year citation data of 2015 CO papers); ・Analyze the co-authorship networks of returnee and non-returnee authors; ・Analyze the pre-2018 publications of returnee and non-returnee authors.	2015-2018	2,000 in each year for total of 8,000
C-ROW papers	・Estimate the quantity of C-ROW papers with returnee authors; ・Estimate the scientific impact of C-ROW papers with and without returnee authors (CiteScore & Citation for 2018 papers); ・Analyze the co-authorship networks of returnee and non-returnee authors; ・Analyze the pre-2018 publications of returnee and non-returnee authors.	2018	2,000
ROW papers (The rest of the world papers)	・As the control group in our citation analysis to indicate the citation behaviors of papers without US and China address; ・The numbers of the rest of the world papers in 2016-2018 are counted based on the query string results (population counts).	2015	2,000

Appendix Table A3. US and China addressed papers with diaspora or returnee author, 2018

		#Papers	Proportion of all China-addressed papers	Proportion of all US-addressed papers
USO with Diaspora author		52,863	-	15.0%
US-ROW with Diaspora author		18,010	-	5.1%
USO without Diaspora author		140,352	-	39.8%
US-ROW without Diaspora author		99,040	-	28.1%
CO with Returnee author		99,685	27.1%	-
C-ROW with Returnee author		19,770	5.4%	-
CO without Returnee author		169,369	46.0%	-
C-ROW without Returnee author		37,286	10.1%	-
US-C-ROW with D	With Returnee authors	2,658	0.7%	0.8%
US-C with D		9,783	2.7%	2.8%
US-C-ROW without D		2,895	0.8%	0.8%
US-C without D		6,240	1.7%	1.8%
US-C-ROW with D	Without Returnee authors	2,635	0.7%	0.7%
US-C with D		8,976	2.4%	2.5%
US-C-ROW without D		3,143	0.9%	0.9%
US-C without D		5,930	1.6%	1.7%

Appendix A4. Estimating the Population of Diaspora Authors in the US, 2018

	Population based	Sample based		Estimated total #AU-IDs = *#papers* × *avg#* AU-ID *per paper*	
	#papers	Avg #US-ND AU-ID per paper	Avg #USD AU-ID per paper	#US-ND AU-ID	#USD AU-ID
USO with D	52,863	3.1	2.1	162,855	112,059
US-C with D	18,759	1.3	1.8	25,151	34,199
US-C-O with D	5,293	3.2	2.6	17,063	13,626
US-other with D	18,010	4.0	1.8	72,293	32,123
USO without D	140,352	5.2	0	729,901	0
US-C without D	12,170	2.2	0	26,636	0
US-C-O without D	6,038	3.8	0	22,733	0
US-other without D	99,040	3.3	0	326,981	0
Total	**352,525**	**3.9**	**0.5**	**1,383,613**	**192,007**

The total #AU-ID exceeds the #authors because one AU-ID could appear more than one time (this is also a measure of the productivity). To estimate the actual number of authors, we need to estimate the #times an AU-ID appears in 2018 papers. Dividing the total #AU-IDs by the average #times an AU-ID appears gives us the estimated author population.

#times an AU-ID appears in 2018 papers (#JA per AU-ID)		
	Sample based	
	US-ND AU-ID	USD AU-ID
USO with D	4.8	3.8
US-C with D	5.9	6.3
US-C-O with D	8.9	6.3
US-other with D	6.1	4.0
USO without D	4.9	-
US-C without D	6.1	-
US-C-O without D	9.6	-
US-other without D	5.6	-
Weighted #JA per AU-ID in 2018	***5.3***	***4.4***

Estimated Author Population in 2018 =
Total AU-IDs / Weighted #JA per AU-ID in 2018

	US-ND author	USD author
Estimated Author Population	***1,383,613/5.3= 261,894 (85.8%)***	***192,007/4.4= 43,184 (14.2%)***

Note: Based on 8,000 sampled US-addressed papers per Appendix Table A1.

Appendix A5. Estimating the Population of Returnee Authors in China, 2018

	Population based	Sample based		Estimated total #AU-IDs = *#papers* × *avg#* AU-ID *per paper*	
	#papers	Avg #China-NR AU-ID per paper	Avg #Returnee AU-ID per paper	#China-NR AU-ID	#Returnee AU-ID
CO	269,054	6.2	0.5	1,659,794	137,621
China-other	57,056	3.9	0.5	223,089	27,273
US-China-other	11,331	3.9	0.8	44,308	9,594
US-China	30,929	4.9	0.8	151,739	25,116
Total	**368,370**	**5.6**	**0.5**	**2,078,930**	**199,604**

The total #AU-ID exceeds the #authors because one AU-ID could appear more than one time (this is also a measure of the productivity). To estimate the actual number of authors, we need to estimate the #times an AU-ID appears in 2018 papers. Dividing the total #AU-IDs by the average #times an AU-ID appears gives us the estimated author population.

#times an AU-ID appears in 2018 papers (#JA per AU-ID)		
	Sample based	
	China-NR AU-ID	Returnee AU-ID
CO	5.8	13.0
China-other	7.4	15.2
US-China-other	8.0	12.0
US-China	5.4	10.8
Weighted #JA per AU-ID in 2018	**6.0**	**13.0**

Estimated Author Population in 2018 =
Total AU-IDs / Weighted #JA per AU-ID in 2018

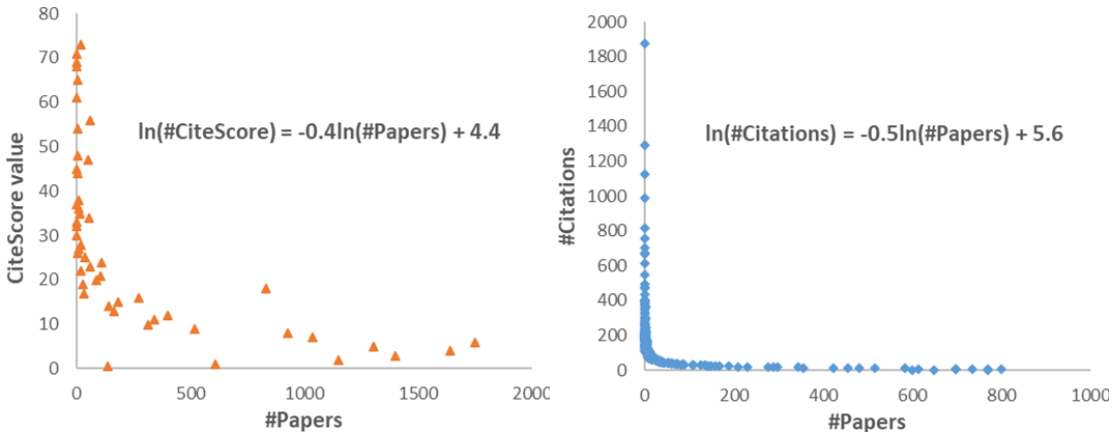
	China-NR author	Returnee author
Estimated Author Population	***2,078,930/6.0= 346,750 (95.8%)***	***199,604/13.0= 15,348 (4.2%)***

Note: Based on 8,000 sampled China-addressed papers per Appendix Table A1 & A2.

Appendix B: Scientific quality/impact of diaspora & returnee research in data set

This appendix documents the results of our analysis of CiteScores and citations, including various "robustness" on the functional form of the relations.

Appendix Figure B1. Power law distributions of CiteScores and 3-year citations of 2018 papers



Note: For simplicity, we use Ln-Ln regression to estimate to powers of each distribution. Power law distributions of CiteScore are plotted based on 13,928 papers with valid author information, address information, and CiteScore values in our dataset. Power law distributions of 3-year citations are plotted based on 14,441 papers with valid author and address information in our dataset. The number of observations is smaller for CiteScores than for citations comes from papers that were published in newly established journals for which Scopus did not yet a CiteScore value.

Appendix Table B1. Average CiteScores and 3-year citations of 2018 US-addressed papers

	CiteScores	3-year citations
US-addressed papers	9.0	16.9
Non-diaspora papers	**8.4**	**14.9**
USO without D	8.3	13.9
US-ROW without D	8.7	15.7
US-C-ROW without D	9.4	24.4
US-C without D	7.1	14.1
Diaspora papers	**10.5**	**22.3**
USO with D	10.5	20.8
US-ROW with D	11.7	25.6
US-C-ROW with D	12.2	36.9
US-C with D	8.7	19.1

Appendix Table B2. Average CiteScores and 3-year citations of 2018 China-addressed papers

	CiteScores		3-year citations	
	Without returnee author	With returnee authors	Without returnee author	With returnee authors
CO	3.4	5.0	11.8	17.1
C-ROW	4.2	5.4	17.4	21.6
US-C-ROW with D	11.9	12.4	33.8	39.8
US-C with D	7.7	9.6	14.8	22.9
US-C-ROW without D	8.2	10.6	21.7	27.4
US-C without D	6.2	7.9	11.5	16.5
China-addressed papers	**3.9**	**5.8**	**13.2**	**18.8**
(Average for all)	(4.6)		(15.3)	

Appendix Table B3. Regression Estimates of Effect of Diaspora and Returnee Papers, Research Field, and Number of Authors of a Paper on CiteScore and Citations of 2018 US/China addressed Papers

US addressed papers						
	CiteScore	***CiteScore***	***CiteScore***	***Citations***	***Citations***	***Citations***
Diaspora paper dummy	2.10*** (0.171)	1.26*** (0.168)	1.94*** (0.055)	7.44*** (0.854)	5.92*** (0.872)	6.66*** (0.848)
#Authors	-		0.05*** (0.004)	-	-	0.28*** (0.019)
Field dummy (21)	NO	Yes	NO	NO	Yes	NO
#Obs	10,127	10,127	10,127	10,441	10,441	10,441
Adjusted R^2	0.0146	0.0819	0.0345	0.0071	0.0177	0.0263
China addressed papers						
	CiteScore	***CiteScore***	***CiteScore***	***Citations***	***Citations***	***Citations***
Returnee paper dummy	2.03*** (0.155)	1.33*** (0.145)	1.64*** (0.149)	6.21*** (0.867)	5.51*** (0.866)	5.20*** (0.860)
#Authors	-	-	0.09*** (0.005)	-	-	0.41*** (0.031)
CO	-	--	-	-	--	-
C-ROW	-	0.68*** (0.201)	-	-	5.21*** (1.192)	-
US-C	-	3.85*** (0.201)	-	-	2.54** (1.201)	-
US-C-ROW	-	6.52*** (0/202)	-	-	15.82*** (1.209)	-
Field dummy (21)	NO	NO	Yes	NO	NO	Yes
#Obs	7,583	7,583	7,583	7,908	7,908	7,908
Adjusted R^2	0.0218	0.1665	0.1185	0.0063	0.0302	0.0379

Note: 95% confidence interval, * p<0.1; ** p<0.05; *** p<0.01. Standard errors are in brackets. Observations without valid address information, author information, or CiteScore value are dropped.

"--" is the benchmark in the regression and "-" means that variable is not added into the regression.

Appendix Table B4. Regression Estimates of Effect of Number of Diaspora and Returnee Authors on CiteScore and Citations of 2018 US and China addressed Papers

	USO&US-ROW papers		CO&C-ROW papers		USC&USC-ROW papers	
	CiteScore	*Citations*	*CiteScore*	*Citations*	*CiteScore*	*Citations*
#Diaspora author	0.41*** (0.067)	2.18*** (0.315)	-	-	0.79*** (0.086)	2.75*** (0.506)
#Non-Diaspora US addressed authors	0.07*** (0.011)	0.37*** (0.051)	-	-	-0.004 (0.0233)	-0.03 (0.139)
#Returnee authors	-	-	0.8*** (0.06)	2.56*** (0.492)	0.49*** (0.112)	1.92*** (0.666)
#Non-retunree China addressed authors	-	-	0.14*** (0.018)	0.99*** (0.146)	-0.03 (0.04)	-0.21 (0.234)
#Rest-of world country authors	0.03*** (0.005)	0.15*** (0.025)	0.17*** (0.014)	1.15*** (0.121)	0.04*** (0.012)	0.37*** (0.071)
USO	--	--	-	-	-	-
US-ROW	0.33 (0.222)	3.83*** (1.052)	-	-	-	-
US-C-ROW	-	-	-	-	2.36*** (0.271)	10.83*** (1.599)
US-C	-	-	-	-	--	--
C-ROW	-	-	0.52*** (0.109)	3.3*** (0.905)	-	-
CO	-	-	--	--	-	-
Field dummy (21)	Yes	Yes	Yes	Yes	Yes	Yes
#Obs	6,345	6,533	3,801	4,000	3,782	3,908
Adjusted R2	0.108	0.0446	0.1973	0.0689	0.1226	0.0519

Note: 95% confidence interval, * p<0.1; ** p<0.05; *** p<0.01. Standard errors are in brackets. Observations without valid address information, author information, or CiteScore value are dropped.

"--" is the benchmark in the regression and "-" means that variable have no meaning in the regression.

Appendix Table B5. Regression Estimates of Effect of Diaspora and Returnee Papers on Ln(CiteScore) and Ln(Citations) of 2018 US/China addressed Papers

	USO&US-ROW papers		**CO&C-ROW papers**		**USC&USC-ROW papers**	
	LN(CiteScore)	*LN(Citations)*	*LN(CiteScore)*	*LN(Citations)*	*LN(CiteScore)*	*LN(Citations)*
Diaspora paper dummy	0.08*** (0.02)	0.16*** (0.029)	-	-	0.1*** (0.024)	0.1*** (0.035)
Returnee paper dummy	-	-	0.23*** (0.026)	0.16*** (0.035)	0.12*** (0.024)	0.16*** (0.035)
Diaspora and Returnee dummy					-0.07 (0.047)	0.12* (0.068)
LN(#Authors)	0.18*** (0.013)	0.41*** (0.02)	0.39*** (0.027)	0.42*** (0.036)	0.26*** (0.02)	0.43*** (0.029)
USO	--	--	-	-	-	-
US-ROW	-0.05*** (0.02)	-0.01 (0.03)	-	-	-	-
US-C-ROW	-	-	-	-	0.18*** (0.025)	0.31*** (0.037)
US-C	-	-	-	-	--	--
C-ROW	-	-	0.18*** (0.025)	0.19*** (0.034)	-	-
CO	-	-	--	--	-	-
Field dummy (21)	Yes	Yes	Yes	Yes	Yes	Yes
#Obs	6,345	6,251	3,801	3,796	3,782	3,794
Adjusted R2	0.2191	0.1593	0.2502	0.0996	0.2036	0.1472

Note: 95% confidence interval, * p<0.1; ** p<0.05; *** p<0.01. Standard errors are in brackets. Observations without valid address information, author information, or CiteScore value are dropped.

"--" is the benchmark in the regression and "-" means that variable have no meaning in the regression.

Appendix Table B6. Regression Estimates of Effect of Number of Diaspora and Returnee Authors on Ln(CiteScore) and Ln(Citations) of 2018 US and China addressed Papers

	USO&US-ROW papers		**CO&C-ROW papers**		**USC&USC-ROW papers**	
	LN(CiteScore)	*LN(Citations)*	*LN(CiteScore)*	*LN(Citations)*	*LN(CiteScore)*	*LN(Citations)*
#Diaspora author	0.047*** (0.0056)	0.107*** (0.0083)	-	-	0.073*** (0.0079)	0.092*** (0.0116)
#Non-Diaspora US addressed authors	0.004*** (0.0009)	0.008*** (0.0013)	-	-	-0.002 (0.0022)	-0.003 (0.0032)
#Returnee authors	-	-	0.19*** (0.016)	0.14*** (0.021)	0.056*** (0.0104)	0.057*** (0.0153)
#Non-retunree China addressed authors	-	-	0.04*** (0.005)	0.04*** (0.006)	-0.003 (0.0037)	0.003 (0.0054)
#Rest-of world country authors	0.002*** (0.0004)	0.004*** (0.0007)	0.03*** (0.004)	0.05*** (0.005)	0.004*** (0.0011)	0.007*** (0.0016)
USO	--	--	-	-	-	-
US-ROW	0.042* (0.0186)	0.217*** (0.0281)	-	-	-	-
US-C-ROW	-	-	-	-	0.249*** (0.0251)	0.442*** (0.037)
US-C	-	-	-	-	--	--
C-ROW	-	-	0.21*** (0.029)	0.18*** (0.039)	-	-
CO	-	-	--	--	-	-
Field dummy (21)	Yes	Yes	Yes	Yes	Yes	Yes
#Obs	6,345	6,251	3,801	3,796	3,782	3,794
Adjusted R2	0.2078	0.1268	0.2365	0.0938	0.1822	0.108

Note: 95% confidence interval, * p<0.1; ** p<0.05; *** p<0.01. Standard errors are in brackets. Observations without valid address information, author information, or CiteScore value are dropped.

"--" is the benchmark in the regression and "-" means that variable have no meaning in the regression.

Appendix Table B7. Regression Estimates of Effect of Diaspora and Returnee Papers on CiteScore and Citations of all US addressed papers and all China addressed Papers, 2018

	US addressed papers				China addressed papers			
	CiteScore	*Citations*	*CiteScore*	*Citations*	*CiteScore*	*Citations*	*CiteScore*	*Citations*
Diaspora paper dummy	1.32*** (0.171)	5.95*** (0.884)	-	-	1.59*** (0.198)	6.65*** (1.219)	-	-
Returnee paper dummy	1.36*** (0.267)	5.63*** (1.376)	-	-	1.04*** (0.14)	4.71*** (0.861)	-	-
#Authors	0.04*** (0.004)	0.23*** (0.019)	-	-	0.06*** (0.005)	0.33*** (0.031)	-	-
#Diaspora author	-	-	0.53*** (0.052)	2.27*** (0.269)	-	-	0.83*** (0.064)	2.89*** (0.398)
#Non-Diaspora US addressed authors	-	-	0.06*** (0.009)	0.32*** (0.049)	-	-	-0.02 (0.017)	-0.12 (0.108)
#Returnee authors	-	-	0.47*** (0.116)	1.71*** (0.599)	-	-	0.52*** (0.068)	1.73*** (0.421)
#Non-retunree China addressed authors	-	-	-0.05 (0.041)	-0.28 (0.212)	-	-	0.01 (0.023)	0.17 (0.14)
#Rest-of world country authors	-	-	0.03*** (0.005)	0.17*** (0.024)	-	-	0.05*** (0.009)	0.42*** (0.054)
USO	--	--	--	--	-	-	-	-
US-ROW	0.27 (0.212)	3.00*** (1.095)	0.46** (0.213)	3.78*** (1.102)	-	-	-	-
US-C-ROW	0.26 (0.279)	9.27*** (1.438)	1.09*** (0.27)	13.34*** (1.393)	5.37*** (0.226)	10.08*** (1.389)	5.56*** (0.226)	11.49*** (1.395)
US-C	-2.05*** (0.278)	-2.77* (1.43)	-1.04*** (0.278)	2.19 (1.434)	2.89*** (0.23)	-1.91 (1.418)	3.09*** (0.212)	0.18 (1.31)
C-ROW	-	-	-	-	0.86*** (0.196)	5.11*** (1.196)	0.76*** (0.202)	4.38*** (1.241)
CO	-	-	-	-	--	--	--	--
Field dummy (21)	yes	yes	yes	yes	yes	yes	yes	yes
#Obs	10,127	10,441	10,127	10,441	7,583	7,908	7,583	7,908
Adjusted R^2	0.1061	0.0452	0.1099	0.0475	0.2333	0.0558	0.2416	0.0551

Note: 95% confidence interval, * p<0.1; ** p<0.05; *** p<0.01. Standard errors are in brackets. Observations without valid address information, author information, or CiteScore value are dropped. "--" is the benchmark in the regression and "-" means that variable have no meaning in the regression. With one exception, the estimated coefficients on our categorization of papers by addresses show that US-China internationally collaborative papers receive higher CiteScores and more citations than other US-addressed papers (where USO is the base group) and receive higher CiteScores and more citations than other China-addressed papers (where CO is the base group). The exception are US-C collaborations, which obtain small negative estimated impacts on CiteScore and citations than USO that we trace mechanically to US-C collaborations having a smaller proportion of US-addressed authors than other US international collaborations. Calculations in Appendix Table B9 shows that conditional on the percentage of US researchers on collaborations and third country researchers, US-C papers obtain modestly higher CiteScore and many more citations than USO papers.

Appendix Table B8. Regression Estimates of Effect of Diaspora and Returnee Papers on Ln(CiteScore) and Ln(Citations) of all US addressed papers and all China addressed Papers, 2018

	US addressed papers				China addressed papers			
	Ln(CiteScore)	*Ln(Citations)*	*Ln(CiteScore)*	*Ln(Citations)*	*Ln(CiteScore)*	*Ln(Citations)*	*Ln(CiteScore)*	*Ln(Citations)*
Diaspora paper dummy	0.09*** (0.015)	0.14*** (0.022)	-	-	0.09*** (0.025)	0.12*** (0.034)	-	-
Returnee paper dummy	0.13*** (0.023)	0.16*** (0.034)	-	-	0.17*** (0.018)	0.15*** (0.024)	-	-
Ln(#Authors)	0.2*** (0.011)	0.42*** (0.016)	-	-	0.34*** (0.016)	0.44*** (0.022)	-	-
#Diaspora author	-	-	0.055*** (0.0046)	0.1*** (0.0067)	-	-	0.07*** (0.008)	0.1*** (0.011)
#Non-Diaspora US addressed authors	-	-	0.003*** (0.0008)	0.007*** (0.0012)	-	-	-0.01** (0.002)	-0.01*** (0.003)
#Returnee authors	-	-	0.051*** (0.0101)	0.045*** (0.0149)	-	-	0.08*** (0.009)	0.07*** (0.012)
#Non-retunree China addressed authors	-	-	-0.004 (0.0036)	0 (0.0053)	-	-	0.01*** (0.003)	0.01*** (0.004)
#Third country authors	-	-	0.002*** (0.0004)	0.004*** (0.0006)	-	-	0.01*** (0.001)	0.01*** (0.002)
USO	--	--	--	--	-	-	-	-
US-ROW	-0.06*** (0.019)	-0.01 (0.028)	0.05*** (0.019)	0.22*** (0.028)	-	-	-	-
US-C-ROW	-0.08*** (0.025)	0.16*** (0.036)	0.1*** (0.024)	0.5*** (0.035)	0.76*** (0.029)	0.31*** (0.04)	0.91*** (0.029)	0.49*** (0.04)
US-C	-0.27*** (0.024)	-0.14*** (0.036)	-0.13*** (0.024)	0.09** (0.036)	0.57*** (0.029)	-0.04 (0.04)	0.65*** (0.027)	0.05 (0.038)
C-ROW	-	-	-	-	0.2*** (0.024)	0.21*** (0.034)	0.24*** (0.026)	0.25*** (0.036)
CO	-	-	-	-	--	--	--	--
Field dummy (21)	yes	yes	yes	yes	yes	yes	yes	yes
#Obs	10127	10045	10127	10045	7583	7590	7583	7590
Adjusted R^2	0.2068	0.158	0.1911	0.1208	0.3458	0.1328	0.3155	0.0992

Note: 95% confidence interval, * p<0.1; ** p<0.05; *** p<0.01. Standard errors are in brackets. Observations without valid address information, author information, or CiteScore value are dropped. The lower numbers of observations for regressions using Ln of dependent variables than using absolute number is due to the observations with 0 citation or 0 CiteScore. "--" is the benchmark in the regression and "-" means that variable have no meaning in the regression.

Appendix Table B9. Regression of CiteScores and Citations of US addressed papers on percent of US addressed authors and percent of third country authors

	CiteScore	3-year citations
Precent of US addressed authors	3.42***(0.532)	9.24***(2.741)
Precent of third country addressed authors	1.13*(0.61)	8.79***(3.143)
#Authors	0.05***(0.004)	0.24***(0.02)
US-ROW	1.21***(0.312)	1.9(1.613)
US-C-ROW	2.71***(0.354)	14.74***(1.828)
US-C	0.84**(0.408)	6.14***(2.101)
Field dummy (21)	Yes	Yes
#Obs	10,142	10,441
Adjusted R^2	0.1036	0.0405

Note: 95% confidence interval, * p<0.1; ** p<0.05; *** p<0.01. Standard errors are in brackets.

Appendix Table B10. Diaspora and returnee papers published on Science and Nature, 2018

	Presence count						Fractional count			
	All articles	Articles with valid name and address info	US addressed articles	US/all	Diaspora articles	Diaspora/US	US addressed author	US/all	Diaspora author	diaspora/US
Science	967	859	581	67.6%	251	43.2%	444.4	51.7%	55.6	12.5%
Nature	1,230	987	658	66.7%	280	42.6%	476.2	48.2%	57.6	12.1%
S&N	2,197	1,846	1,239	67.1%	531	42.9%	920.6	49.9%	113.2	12.3%
	All articles	Articles with valid name and address info	China addressed articles	China/all	Returnee articles	Returnee/China	China addressed author	China/all	Returnee author	returnee/China
Science	967	859	115	13.4%	87	75.65%	55.4	6.4%	15.7	28.3%
Nature	1,230	987	136	13.8%	103	75.74%	51.9	5.3%	14.8	28.6%
S&N	2,197	1,846	251	13.6%	190	75.70%	107.3	5.8%	30.5	28.5%

Note: The difference between column 1 and 2 is number of Scopus journal article records without author name and address information. By checking those papers on Science and Nature websites, we found that those papers are non-research articles (news, comments, etc.) but mis-recorded as research articles in Scopus. We exclude those papers from our analysis

Appendix C: Diaspora and returnee authors in US-China collaboration

This appendix details how we used fractional counts to create the pooled samples for testing whether the observed distribution of diaspora and returnee authors on US-China collaborations could have come from random selection from pooled samples of possible authors.

Method of fractional count calculation

We created a pool of US-addressed authors and a pool of China-addressed authors from the 2018 papers based on a fractional count of Author IDs for Table 2 analysis. The fractional count adjusts the number of authors in the pool upwards by the number of papers they wrote in 2018 and downwards by the number of co-authors on papers. For example, if a paper has one diaspora author, one US addressed non-diaspora author and one UK author, we count 1/3 for diaspora author and 2/3 for all US addressed authors, then the diaspora share of fractionated US addressed author equals one-third divided by two-thirds = 1/2. As shown in Appendix Table C2 and C3, the diaspora share of fractionated US-addressed authors is 12%. The returnee share of fractionated China-addressed authors is 8.5% in 2018.

Appendix Table C1. Proportions of types of authors on US-addressed papers, 2018

	%US-addressed and non-Chinese named author	%Diaspora author	%China-addressed author	% ROW-addressed author
USO without D	100.0%	0.0%	0.0%	0.0%
US-ROW without D	36.5%	0.0%	0.0%	63.5%
US-C-ROW without D	25.6%	0.0%	35.0%	39.4%
US-C without D	29.8%	0.0%	70.2%	0.0%
USO with D	59.2%	*40.8%*	0.0%	0.0%
US-ROW with D	43.4%	19.3%	0.0%	37.3%
US-C-ROW with D	21.9%	17.5%	34.8%	25.9%
US-C with D	18.2%	24.8%	57.0%	0.0%

Appendix Table C2. Fractional count of US-addressed papers by types of authors, 2018

	#Papers	a. Fractional count by % US-addressed and non-Chinese named author	b. Fractional count by % Diaspora author	c. Fractional count by % China-addressed author	d. Fractional count by % ROW-addressed author	***e. Fractional by % all US-addressed author (a+b)***	***b/e***
US-addressed papers	352,525	225,424 (63.9%)	30,600 (8.7%)	23,186 (6.6%)	73,316 (20.8%)	**256,024**	**12.0%**
USO without D	140,352	140,352	0	0	0	**140,352**	**0.0%**
US-ROW without D	99,040	36,190	0	0	62,850	**36,190**	**0.0%**
US-C-ROW without D	6,038	1,544	0	2,115	2,379	**1,544**	**0.0%**
US-C without D	12,170	3,624	0	8,546	0	**3,624**	**0.0%**
USO with D	*52,863*	31,315	*21,548*	0	0	**52,863**	**40.8%**
US-ROW with D	18,010	7,818	3,474	0	6,718	**11,291**	**30.8%**
US-C-ROW with D	5,293	1,159	925	1,840	1,369	**2,084**	**44.4%**
US-C with D	18,759	3,422	4,653	10,684	0	**8,075**	**57.6%**

Note: The fractional count of papers is calculated by multiplying the #papers in the first column by percent of authors on the same types of papers shown in Appendix Table C1. For example, the fractional count of USO with D by diaspora authors = 52,863 * 40.8% from Appendix Table C1 = 21,548

Appendix Table C3. Fractional count of China-addressed papers by returnee-diaspora authors, 2018

	#Papers	a. Fractional count of papers by China-addressed authors	b. Fractional count of papers by returnee authors	**b/a**
China-addressed papers	368,370	325,572	27,692	**8.5%**
CO	269,054	269,054	22,142	**8.2%**
C-ROW	57,056	33,332	4,022	**12.1%**
US-C-ROW with D	5,293	1,840	204	**11.1%**
US-C with D	18,759	10,684	670	**6.3%**
US-C-ROW without D	6,038	2,115	226	**10.7%**
US-C without D	12,170	8,546	428	**5.0%**

We randomly sampled 2,000 diaspora authors & 2,000 non-diaspora authors from the US addressed papers described in Appendix Table A1 and 2,000 returnee authors & 2,000 non-returnee authors from the China addressed papers described in Appendix Table A1 and A2. We retrieved English journal articles published in 2018 for 8,000 randomly chosen authors from the two (which goes beyond the papers in our initial sample) and calculated the total number of papers each wrote in 2018 and the number that were US-China collaborations. About the heterogeneity in authors' propensity to work on US-C collaborations. While some of the variation reflects the number of papers that authors wrote in 2018, as the proportion of collaborations for an author with one paper is necessarily 0% or 100%, and so on, there is greater variation in the US-China collaboration proportion of papers among authors with the same numbers of papers than would occur if each had the same propensity to work on a US-China collaboration.

Appendix Table C4. Sampled Diaspora, Non-Diaspora, Returnee, and Non-Returnee authors for Table 3 Analysis

Author type	#Sampled Author IDs	#Valid Author IDs at 2022	#English journal articles by authors in 2018
Diaspora	2,000	1,972	8,987
Non-Diaspora	2,000	1,966	9,889
Returnee	2,000	1,974	22,836
Non-Returnee	2,000	1,936	11,485

Note: 1.8% of Author IDs sampled in early 2020 became invalid in 2022 because Scopus updates Author ID over time.

Appendix D: Author career and citation network

Appendix Table D1. Estimated number of US-China dual addressed authors in 2018

	Fractional count of authors	Fractional share of US addressed author
Dual diaspora author	*7,264*	*0.461%*
Dual non-diaspora author	*1,232*	*0.078%*
Non-dual diaspora author	181,811	11.5%
Non-dual non-diaspora author	1,385,313	87.9%
	Fractional count of authors	**Fractional share of China addressed authors**
Dual returnee author	*4,273*	*0.188%*
Dual non-returnee author	*150*	*0.007%*
Non-dual returnee author	189,402	8.3%
Non-dual non-returnee author	2,084,709	91.5%
Total dual	***12,919***	**-**

Note: US-China dual addressed authors are count as part of US addressed author and part of China addressed authors based on their affiliation address, if one dual author listed one US affiliation, one China affiliation and one UK affiliation, this dual author will be count as 1/3 US addressed author and 1/3 of China addressed author. Estimated number of US-China dual addressed authors is based sampled 2,000 USC and 2,000 US-C-ROW papers per Appendix Table A1. Estimated number of non-dual US and China addressed authors is based on 12,000 sampled US and China addressed papers per Appendix Table A1 and A2.

In Appendix A4, we estimate the number of US addressed authors on 2018 US-China collaborative papers is 139,408 (authors will be count n times if their AU-IDs appeared on n papers), so dual authors' share of US addressed authors on collaborative papers = (7264+1232)/139408 = 6.1%. Similarly, the estimated number of China addressed authors on 2018 US-China collaborative papers is 230,757 (see Appendix A5) and the dual authors' share of China addressed authors on collaborative papers = (4273+150)/230757 = 1.9%.

Appendix Table D2. Pre-2018 publications of dual addressed authors

Dual authors with at least one pre-2018 publication where the author had	Pct of authors
1. A dual address	71%
2. A solo China address	89%
3. A solo US address	61%
Union of the 1-3 columns	100%
4. A solo US address and a China addressed coauthor	55%
5. A solo China address and a US addressed coauthor	63%
Union of the 4-5 columns	79%

Note: Estimated number of US-China dual addressed authors is based on the sampled 2,000 US-C and 2,000 US-C-ROW papers per Appendix Table A1.

Appendix Table D3. Percent of Pre-2018 Publications by 2018 US-Addressed and China-Addressed Authors, by Diaspora and Returnee Status

Diaspora and returnee status	% of pre-2018 publications by 2018 US-addressed authors where the author had				
	Solo-China address and no US addressed coauthors	Solo-China address and US addressed coauthors	US-China dual address	Solo-US address and no China addressed coauthors	Solo-US address and China addressed coauthors
Diaspora	8.6%	1.4%	67.4%	17.3%	2.4%
Non-D	0.1%	0.1%	86.6%	3.7%	0.1%
Returnee	68.6%	10.5%	6.7%	1.9%	2.9%
Non-R	77.6%	4.9%	0.0%	0.0%	0.0%

Note: Based on all pre-2018 publications by randomly sampled 2000 diaspora and 2000 non-diaspora authors from the 8000 sampled US-addressed papers described in Appendix Table A1, and 2000 returnee authors and 2000 non-returnee authors from the 8000 sampled China-addressed papers per Appendix Table A1 and A2. In these calculations, to focus on US-China connections, we treat papers where author had China and Rest of World addresses and no US address as "solo China" and papers where author had US and Rest-of-World address and no China address as "solo US address" and treat a paper where author had US, China, and Rest-of-World addresses as US-China dual address.

Appendix Table D4. Citation preference measures between US and China

2015 papers	Citations to 2015 papers from 2016-18 papers		
	from USO D	from USO ND	***DD Ratio***
CO	26,546	27,467	-
ROW	145,823	263,932	-
from D/from ND	***0.182***	***0.104***	***1.75***
	from CO	from ROW	***DD Ratio***
USO D	99,660	185,690	-
USO ND	136,023	516,658	-
from CO/from ROW	***0.733***	***0.359***	***2.04***
	from CO R	from CO NR	***DD Ratio***
USO	107,587	128,096	-
ROW	263,961	530,562	-
from R/from NR	***0.408***	***0.241***	***1.69***
	from USO	from ROW	***DD Ratio***
CO R	28,790	123,472	-
CO NR	25,223	217,526	-
from USO/from ROW	***1.141***	***0.568***	***2.01***

Note: To estimate the diaspora preference to CO vs ROW compared to non-diaspora papers, we randomly sampled 2,000 CO and 2,000 ROW papers published in 2015, then we retrieved 27,443 papers in 2016-2018 that cite our sampled 2015 CO papers and ROW papers. We select the 2016-18 USO papers from those 27,443 papers, and use the same name-based method to calculate the average citations per paper from 2016-18 USO diaspora/non-diaspora papers for our sampled 2015 CO and ROW papers. Multiplying number of CO papers and ROW papers in 2015 (CO: 203,602; ROW: 894,849), we estimated the number of citations from all 2016-2018 diaspora and non-diaspora USO papers to all 2015 CO and ROW papers, respectively.

To estimate the CO preference to diaspora vs non-diaspora compared to ROW, we randomly sampled 2,000 USO papers published in 2015 and use our name-based method to divide the USO papers into 2015 diaspora USO papers and non-diaspora USO papers. Then we retrieved 35,108 papers in 2016-18 that cite our sampled 2015 USO papers. We select the 2016-18 CO and ROW papers from those 35,108 papers and calculate the average citations per paper from 2016-18 CO/ROW papers for our sampled 2015 diaspora and non-diaspora USO papers. Multiplying

number of diaspora and non-diaspora USO papers in 2015 (diaspora USO: 48,629; non-diaspora USO: 147,661), we estimated the number of citations from all 2016-2018 CO and ROW papers to all 2015 diaspora and non-diaspora USO papers, respectively.

To estimate the returnee preference to USO vs ROW compared to non-returnee papers, we use the randomly sampled 2,000 USO and 2,000 ROW papers in 2015, and retrieved 47,883 papers in 2016-2018 that cite our sampled 2015 USO papers and ROW papers. We select the 2016-18 CO papers from those 47,883 papers, and use the same publication-history-based method to calculate the average citations per paper from 2016-18 CO returnee/non-returnee papers for our sampled 2015 USO and ROW papers. Multiplying number of USO papers and ROW papers in 2015 (USO: 196,290; ROW: 894,849), we estimated the number of citations from all 2016-2018 returnee and non-returnee CO papers to all 2015 USO and ROW papers, respectively.

To estimate the USO preference to returnee vs non-returnee compared to ROW, we use the randomly sampled 2,000 CO papers in 2015 and apply our publication-history-based method to divide the CO papers into 2015 returnee CO papers and non-returnee CO papers. Then we retrieved 14,669 papers in 2016-18 that cite our sampled 2015 CO papers. We select the 2016-18 USO and ROW papers from those 14,669 papers and calculate the average citations per paper from 2016-18 USO/ROW papers for our sampled 2015 returnee and non-returnee CO papers. Multiplying number of returnee and non-returnee CO papers in 2015 (returnee CO: 56,935; non-returnee CO: 146,667), we estimated the number of citations from all 2016-2018 USO and ROW papers to all 2015 returnee and non-returnee CO papers, respectively.

Appendix E: Numbers of US F1 and J1 Visas Issued per month to Chinese citizens

Appendix Table E1 shows that the numbers of F1 visas issued to Chinese citizens between January to April 2022 exceed numbers in the same months of 2021, and the monthly F1 visa issuances started declining in May 2022 after the Shanghai lockdown and the US consulate in Shanghai was closing to public in April 2022.

Appendix Table E1. Number and percentage of F1 visas issued by the US to Chinese citizens

a. #F1 visa issued to Chinese citizens by the US						
	2018	**2019**	**2020**	**2021**	**2022**	**2023**
Jan	3,094	3,001	2,918	**480**	**936**	2,356
Feb	925	664	112	**231**	**554**	1,282
March	1,359	1,443	135	**304**	**585**	1,775
Apr	3,700	4,281	17	**321**	**1,550**	-
May	18,271	20,871	7	***23,066***	***8,570***	-
June	29,555	34,001	8	33,896	18,860	-
July	24,666	21,781	145	21,163	15,090	-
Aug	7,118	824	363	8,699	4,909	-
Sept	1,351	1,231	268	1,355	980	-
Oct	922	1,150	217	1,604	948	-
Nov	1,948	2,335	195	4,349	1,983	-
Dec	4,774	7,002	468	3,963	2,546	-
Total	97,683	98,584	4,853	99,431	57,511	5,413
b. %F1 visa issued to Chinese citizens by the US						
	2018	**2019**	**2020**	**2021**	**2022**	**2023**
Jan	17%	17%	19%	**5%**	**7%**	16%
Feb	13%	10%	2%	**6%**	**8%**	19%
March	16%	18%	4%	**6%**	**7%**	18%
Apr	27%	30%	17%	**4%**	**12%**	-
May	42%	45%	3%	***59%***	***23%***	-
June	38%	43%	1%	43%	20%	-
July	28%	25%	1%	22%	16%	-
Aug	17%	2%	2%	14%	9%	-
Sept	14%	15%	4%	11%	10%	-
Oct	13%	15%	3%	13%	11%	-
Nov	14%	17%	2%	16%	8%	-
Dec	18%	23%	2%	11%	6%	-
Total	27%	27%	5%	25%	14%	17%

Note: The numbers of F1 visa issuances to Chinese nationality are from the Monthly Nonimmigrant Visa Issuance Statistics by U.S. Bureau of Consular Affairs (https://travel.state.gov/content/travel/en/legal/visa-law0/visa-statistics/nonimmigrant-visa-statistics/monthly-nonimmigrant-visa-issuances.html).

Appendix Table E2. Number and percentage of J1 visas issued by the US to Chinese citizens

a. #J1 visa issued to Chinese citizens by the US						
	2018	**2019**	**2020**	**2021**	**2022**	**2023**
Jan	2,021	2,114	1,421	33	300	605
Feb	1,251	1,218	105	70	318	499
March	2,476	2,325	71	80	323	515
Apr	3,561	4,722	18	82	284	-
May	4,402	4,503	11	133	416	-
June	4,526	5,732	8	458	934	-
July	5,746	4,389	12	966	989	-
Aug	5,091	4,713	45	647	881	-
Sept	2,812	2,836	60	419	601	-
Oct	2,365	2,113	54	334	446	-
Nov	2,298	1,809	55	728	679	-
Dec	2,560	2,693	65	726	678	-
Total	39,109	39,167	1,925	4,676	6,849	1,619
b. %J1 visa issued to Chinese citizens by the US						
	2018	**2019**	**2020**	**2021**	**2022**	**2023**
Jan	14%	15%	10%	1%	3%	4%
Feb	10%	9%	1%	2%	2%	3%
March	8%	7%	1%	1%	1%	1%
Apr	7%	8%	16%	1%	1%	-
May	7%	7%	2%	1%	1%	-
June	11%	14%	1%	2%	3%	-
July	14%	11%	1%	4%	3%	-
Aug	21%	19%	1%	4%	4%	-
Sept	23%	22%	3%	5%	5%	-
Oct	15%	14%	3%	4%	3%	-
Nov	11%	9%	3%	4%	3%	-
Dec	13%	13%	1%	4%	4%	-
Total	11%	11%	3%	3%	2%	2%

Note: The numbers of J1 visa issuances to Chinese nationality are from the Monthly Nonimmigrant Visa Issuance Statistics by U.S. Bureau of Consular Affairs (https://travel.state.gov/content/travel/en/legal/visa-law0/visa-statistics/nonimmigrant-visa-statistics/monthly-nonimmigrant-visa-issuances.html).