A Stress Test of the United States' Knowledge Supply Chain

Christopher Esposito

UCLA

Abstract

Like physical products, new technologies are created by combining inputs from across the world. This paper identifies the global supply chain of scientific knowledge used by U.S. inventors and tests its resilience to international shocks. Using simulations that remove countries from the supply chain and allow the network to reconfigure, I show that the supply of science relied on by U.S. inventors is resilient to shocks affecting individual countries, but that shocks that affect multiple countries can lead the supply chain to collapse. Network collapse not only impedes the ability for inventors in the U.S. to access foreign science, but also impacts their ability to capitalize on domestic investments in basic research: preliminary estimates suggest that excluding international science from the U.S. economy would decrease the conversion of NSF-funded science into patented technologies by 47% and would slow the speed of conversion by 25%. Furthermore, the exposure of U.S. invention to international disruptions is distributed unevenly across countries and legislative entities: while losing access to Canadian science would reduce the conversion rate of NSF-funded science to patents by 3% and the slow the process by 1%, losing access to science from the European Union would reduce the rate by 25% and slow the process by 15%.

1) Introduction

International systems, by virtue of their scale and the institutional boundaries that they cross, can increase growth and resilience, but their stewardship also requires coordination across jurisdictional boundaries. Some international systems, such as the global supply chains for products, generate mutual benefits that align policymaker incentives across international borders, reducing the likelihood of perverse interventions. To act to in accordance with these incentives, however, policymakers need information on the benefits that an international system generates for their constituents.

In this paper, I measure the contribution of international science to innovation in the United States. Like physical products, inventors create new technologies by assembling knowledge-based inputs into functioning devices (Fleming and Sorenson 2001). Because highly-specific knowledge is required to create new technologies (Jones 2009), many of the knowledge inputs for inventions are found outside the country where a new technology is developed (van der Wouden 2020; Esposito 2023; Wang, Li, and Wang 2024; Wu, Esposito, and Evans 2024). While the global interdependence of innovation is a well-known fact, there have been few attempts to measure its level. This lapse has arisen due to two main reasons.

The first is that the knowledge-based antecedents of new innovations often extend far upstream from a focal technology, into areas of basic research. Several datasets provide records of the first-order linkages between new technologies and knowledge inputs. For example, direct citations from patent documents to scientific articles indicate science relied on by the patent.(Marx and Fuegi 2020) However, using available data, it is more difficult to observe the indirect linkages between science and technology, such as basic research that influences technology through applied research.

The second reason is that, in order to determine a country's dependence on global knowledge, one must not only observe the frequency by which the inventors in a country build on knowledge found in others, but also the availability of substitute knowledge. Without identifying potential substitutes, studies may overstate the dependence of innovation in one country on the science of another.

I address the first issue in this study by tracing the knowledge embedded in patents far back to its antecedents in basic research, using a method pioneered by Ahmadpoor and Jones (2017). I extend this method by tracing the knowledge trajectories of patents all the way back to scientific articles that were supported by funding from the United States' National Science Foundation (NSF). This extension allows me to link investments in basic science to their downstream technological inventions. Using information on where the authors of the papers in these are located, I am able to see how science moves across international borders from the moment that it receives public investment to when it stimulates invention.

I address the second issue – that one must observe counterfactual knowledge supply chains to determine the importance of international science – by generating simulated supply chain networks where specific countries are removed from the graph. These simulated networks not only explore the potential consequences of (extreme) policy interventions, but also measure the difficulty of substituting for the science produced by each country. To create the simulated networks, I re-generate the shortest-path trajectories that link patents to NSF-supported papers, while coercing the trajectories to avoid papers produced in countries of interest. This method uncovers the alternative shortest-path trajectories that connects patents to their knowledge antecedents, and thus generates a counterfactual for how the network might readapt following the removal of certain countries.

My results show that the U.S. knowledge supply chain is robust to the removal of individual countries from the network, but that the removal of multiple countries can cause the network to collapse. For example, when U.S. scientists and inventors lack access to science produced by international scientists, between 47% and 63% of the trajectories that link NSF-supported articles and downstream patents are severed and cannot be replaced by any alternative path. Of the paths that are not severed or are replaceable, the average length of the paths increases by 25%. Moreover, network collapse is a superlinear function of the number of countries (or more precisely, the number of articles) removed from the network, because when only one country or a few articles are removed, many alternative citation paths are available. As more countries and articles are removed, those alternatives dwindle, and the network disintegrates.

2) Data and Methods

2A) Data

My analysis combines patent records with data on scientific publications. The patent data comes from three sources: from the USPTO (PatentsView), I source patent numbers and inventor location information, from the Reliance on Science dataset (Marx and Fuegi 2020), I source citation records on patents made to scientific articles, and from PatStat I source patent family information. I use the patent family records to identify patents that are applied with the three major patent offices (the USPTO, JPO, and EPO), and I restrict my study to these "triadic" patents in order to reduce home-market biases. Furthermore, to reduce potential volatility in the dataset due to patents that are weakly connected to science, I restrict the patent sample to patents that cite 5 or more scientific articles. This results in a sample of 301,237 patents.

For data on scientific publications, I retrieve data from two sources. First, I retrieve paper numbers and citation records from OpenAlex. Second, I source author affiliation information (including location information) from SciSciNet, a datalake that cleans and combines publication information from a variety of sources, particularly OpenAlex. I retrieved affiliation information from SciSciNet rather than directly through OpenAlex because the authors of the dataset have taken measures to improve the quality of the data (Lin, Liu, and Wang 2023). Finally, from SciSciNet I source NSF grant records, which are available at the paper level based on the acknowledgement sections in publications.

2B) Methods

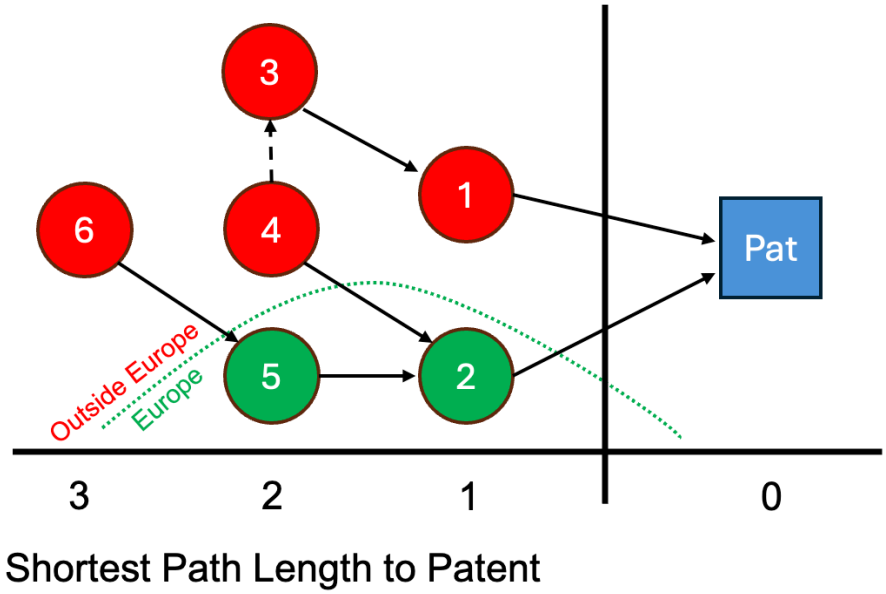
I define knowledge supply chains as the scientific knowledge that influences patents. To identify this knowledge, I adopt the method of Ahmadpoor and Jones (2016) to trace the shortest citation path between patents and scientific articles.

These methods are illustrated in Figure 1. I begin with USPTO records of triadic patents and, using the Reliance on Science dataset (Marx and Fuegi 2020), identify the scientific papers that they cite directly. These direct citations are reflected in Figure 1 by the blue square patent, which receives two inward-pointing edges from two papers, numbered Paper 1 and Paper 2. Because Papers 1 and 2 are directly cited by a patent, they are positioned at a value of 1 on the *x* axis, indicating that the shortest path length from these papers to the nearest patent is 1.

After identifying the papers that are directly cited by patents, I identify the papers that are *indirectly* cited by patents. To do so, I begin with the papers that are directly cited (Papers 1 and 2) and record the papers that they cite: Papers 3, 4, and 5. These papers are located at a value of 2 on the *x* axis, because the shortest length path from them to the nearest patent is 2.

Next, I identify the papers at shortest length distance of 3 from the nearest patent. To do so, I extract the citations made by papers at Distance = 2. In Figure 1, there are two papers at D=2 that cite other articles: Paper 3 cites Paper 4, and Paper 5 cites Paper 6. However, because Paper 4 was also cited by Paper 2, and because Paper 2 is closer to the nearest patent than is Paper 3, the citation that runs from Paper 4 to Paper 3 does not fall on a shortest path to patents. Therefore, I omit this citation from the dataset; it is indicated by a dashed line in Figure 1 to denote its omission.

Figure 1: Methods for Constructing Patent Knowledge Supply Chains Using Patent and Paper Citations



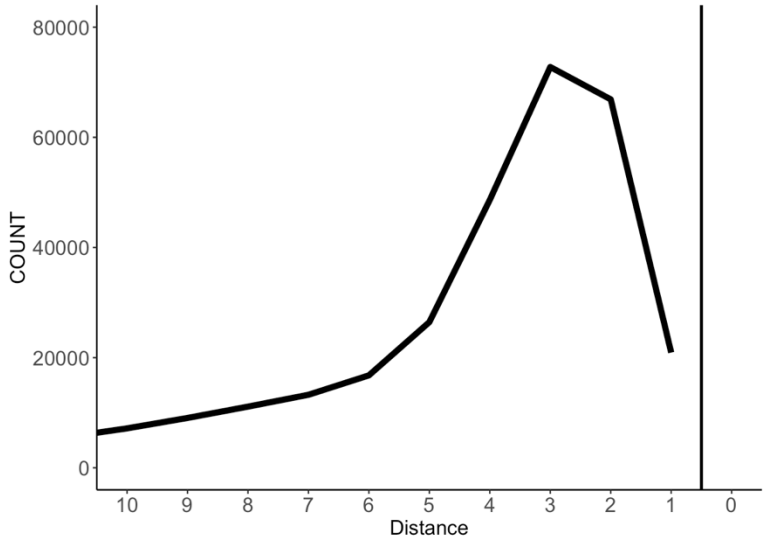
I repeat these methods until I arrive at scientific papers that are a distance of 10 from the nearest patent. This method produces a directed acyclic network with 47 million edges. The network consists of 301,237 patents and 19,478,979 scientific publications.

After constructing the shortest-path network, I extract the paths of the network that begin with papers supported by the NSF. To do so, I conduct walks along the network, beginning with

each paper that acknowledges receipt of NSF support and moving forward in the network across papers until I reach a patent. These walks represent the shortest paths from each NSF-supported article to the nearest patent.

In Figure 2, I plot the number of papers by each value of Distance. The patents are located at D=0, indicating that they are a distance of 0 from the nearest science-citing patent. The distribution has a mode of 3, and a long-left tail.

Figure 2: Distribution of Papers Citing NSF Support by Distance to Nearest Science-Citing Patent



Finally, to test the resilience of the U.S. knowledge supply chain, I generate different versions of the shortest-path network that explore potential scenarios in which certain countries or supranational legislative bodies (such as the EU) might be removed from the network. Particularly, I generate versions of the shortest-path network that exclude science produced in Canada, in China, in the EU, in Europe (the EU plus the U.K., Norway, and Switzerland), and all other countries

outside the U.S. I define science as being produced in a specific country or region if a majority of its authors are affiliated with research institutions that region.

In Figure 1, I illustrated how the shortest path network is constructed after excluding a region from the network. In the figure, Papers 5 and 2 are produced in Europe. These papers are excluded from the version of the network that excludes European science. The effect of removing these papers propagates across the network, because the shortest path of Paper 4 to the patent is through a European paper (Paper 2). After removing Paper 2, the shortest path for Paper 4 runs through Paper 3. This has the effect of moving Paper 4 further away from the patent; its distance to the patent increases from 2 to 3 once Paper 2 is removed from the network.

Figure 1 also demonstrates how the removal of a region can cause some papers to be disconnected from patents. For example, Paper 6 is only cited by European science. If European science is removed from the network, Paper 6 will lose its connection to the network.

Finally, it should be noted that by defining science as produced in a country or region based on whether a majority of the authors are based in that region, I allow excluded regions to still contribute some science to U.S. innovation. These remaining contributions can be interpreted in two ways. First, one might imagine that most policies or interventions that separate the science of countries will be incomplete, and thus that some knowledge will remain transmitted between the two countries. Second, one might imagine that a country that loses access to a specific area of research might try to recreate that specific research, instead of researching "around" the area. For example, if the U.S. were to lose access to Paper 2 in Figure 1, its might try to recreate Paper 2, instead of navigating around Paper 2 along the Paper 4 -> Paper 3 -> Paper 1 -> Patent path. By only excluding papers from regions if a majority of the authors are from that region, I de facto

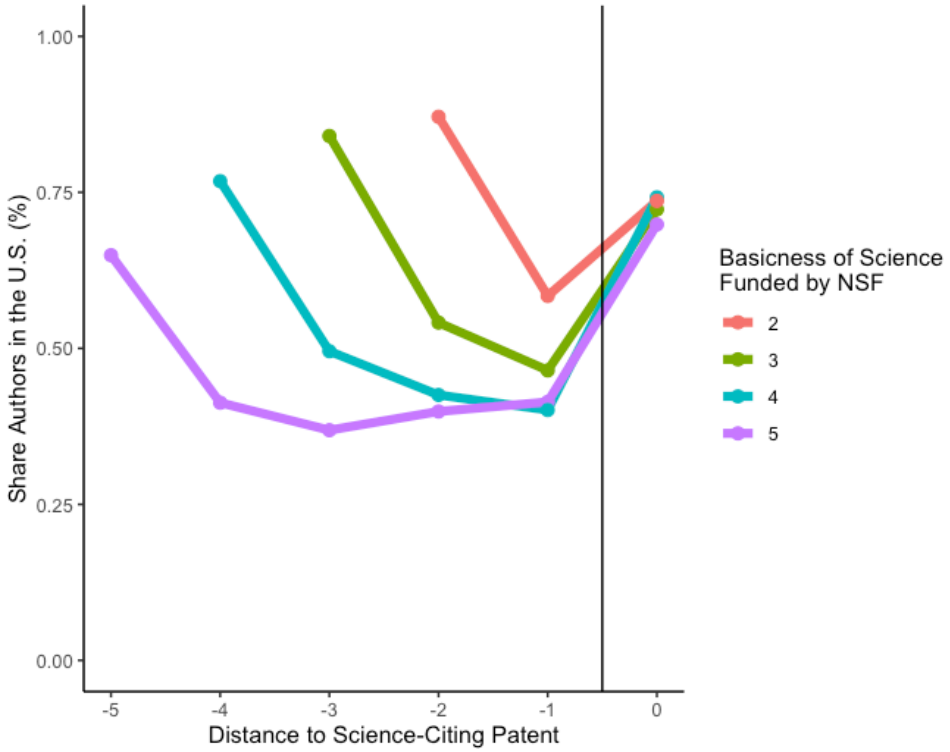
assume that authors from non-excluded regions are able to complete their papers even if a non-majority of their coauthors are removed from the team.

3) Descriptive Analysis

3A) Empirical Background

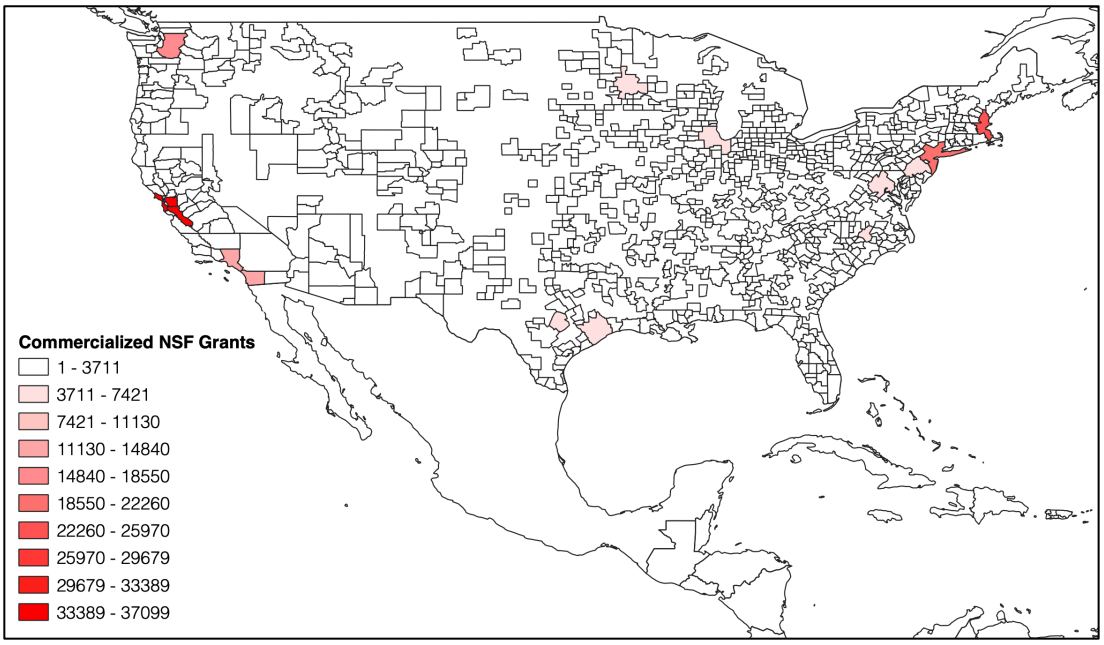
The United States innovation ecosystem makes extensive use of international science. Figure 3 plots the share of authors on papers and patents with U.S.-based affiliations, for papers that fall along walks along the shortest-path citation graph beginning with NSF-supported scientific articles and ending with science-citing patents. The data are grouped based on the distance to the patent boundary of the scientific article supported by the NSF. For each group, the share of authors is high for the paper supported by the NSF, subsequently declines, and jumps back up as the network crosses the patent-science boundary. For example, on NSF-supported articles that are 4 citations removed from the nearest patent (the teal line), 77% of the authors are based in the U.S. As this knowledge becomes more applied, the share of authors in the U.S. declines, until it bottoms out at 40% at a distance of 1 from the patent boundary. However, once the knowledge crosses the patent boundary, the share of U.S.-based authors (inventors) jumps up to 73%.

Figure 3: Share of Authors in the U.S. by Distance to the Nearest Patent



The U.S. is able to leverage global science and capture the downstream patents because it has strong complementary assets for technology capture. The nature of these complementary assets is indicated by the regions of the U.S. that produce patents that build on NSF-supported science. Figure 4 presents a map of U.S. metropolitan areas by the number of patents that they produce that are linked to NSF-supported papers. There are very strong concentrations in the metropolitan areas centered on San Francisco and San Jose, and weaker concentrations in the metro areas centered on New York City, Boston, Seattle, Los Angeles, and San Diego. All other metro areas produce comparatively few NSF-influenced patents. Moreover, the geography of NSF-influenced patenting is similar to the geography of venture capital, suggesting that the United States is able to translate globally-sourced science into patents because it has agglomerations that efficiently match scientific knowledge to financial resources (Samila and Sorenson 2010; 2011).

Figure 4: Number of Patents Linked to NSF-Supported Papers by Metropolitan Area



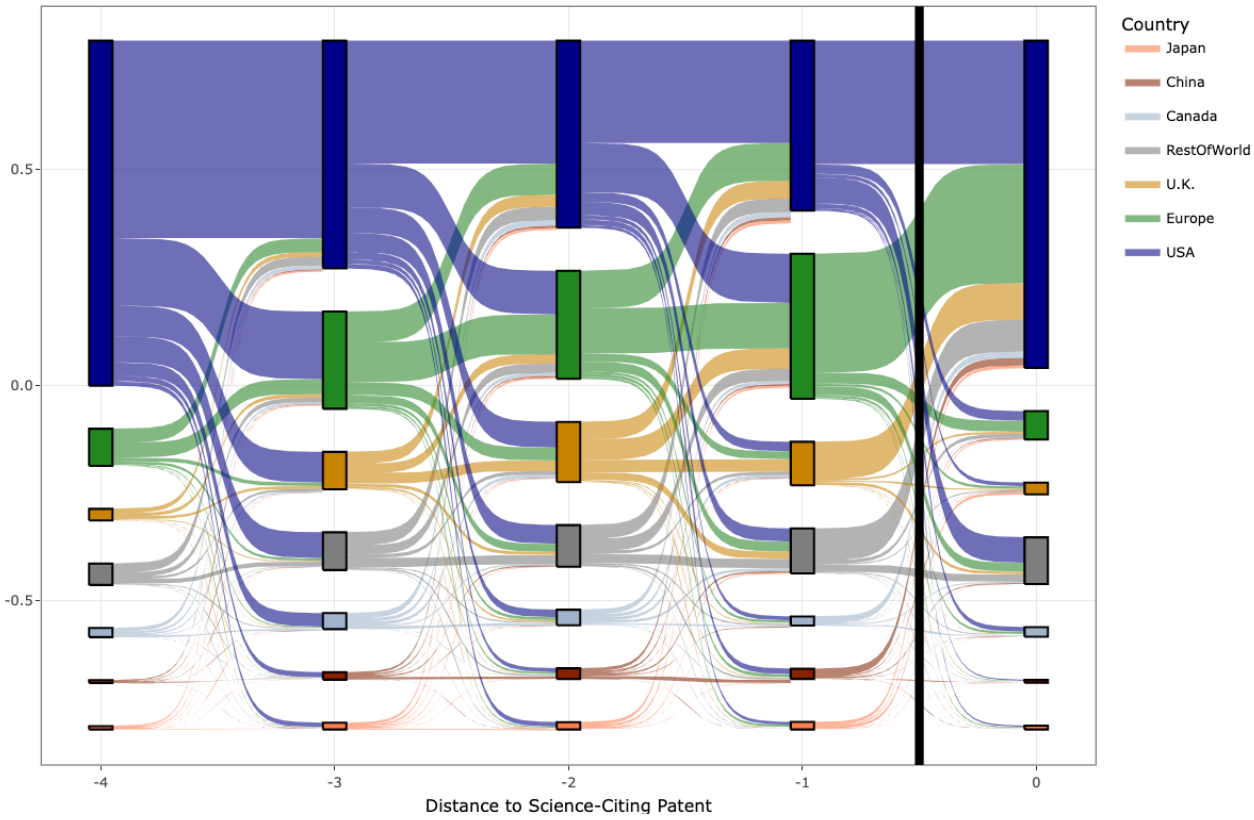
3B) The U.S. Global Supply Chain of Science

Which countries does knowledge pass through as it advances from NSF-supported science to the patenting boundary? Figure 5 presents a Sankey diagram of the knowledge that falls along the knowledge trajectories that begin with NSF-funded articles a distance of 4 citations from the patent boundary. These trajectories are the same as those represented by the turquoise line in Figure 3. In the figure, Europe (the EU27 plus Norway and Sweden) is aggregated to a single entity, and all non-listed countries are aggregated to a "Rest of World" category.

Figure 5 shows that the papers directly supported by the NSF (at Distance = 4) are disproportionately produced by U.S. based inventors, their trajectories pass through many countries as they move toward the patent boundary. In particular, as the trajectories move from D=4 to D=3, many of them move to Europe, the U.K., Canada, and the rest of the world. This process continues through D=1. Once the trajectories cross the patent boundary, much of the

knowledge moves back from Europe, the U.K., Canada, and the rest of the world and is patented in the U.S.

Figure 5: Knowledge Trajectories Linked to NSF-Supported Science at D=4 as they Evolve Toward the Patent Boundary



4) Counterfactual Analyses

4A) Sankey Diagrams of Shortest Path Networks Excluding the EU and the Rest of Europe

To test the resilience of the U.S. knowledge supply chain to shocks external shocks, it is necessary to know not only the countries through which the supply chain passes, but also the ability for knowledge to pass through other countries along other paths. Figure 1 demonstrated how knowledge can pass through other paths when one country or region is removed from the

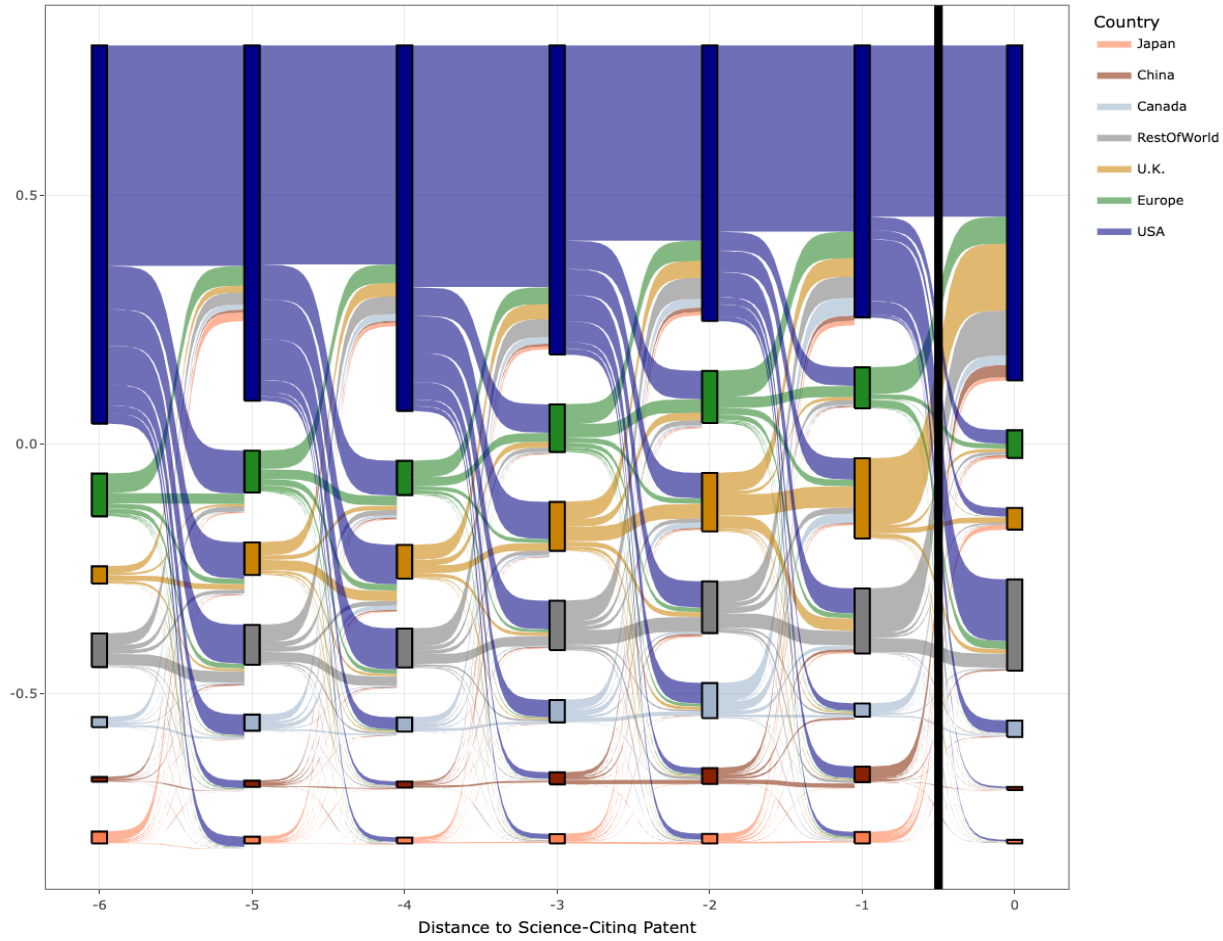
U.S. supply chain. For example, if Europe is removed from the network in Figure 1, a trajectory that starts at Paper 4 could still reach the patent, albeit along a path of longer length.

To assess the supply chain's resilience, I reconstruct the shortest-length path network between NSF-supported articles and patents after excluding specific countries or regions from the network. These alternative networks allow me to assess how the connectivity of NSF-supported papers to patents and the average distance of those connections changes after excluding countries from the network.

Figure 6 illustrates this method by presenting a Sankey diagram of the shortest-path network after excluding papers where the majority of the authors are in the EU. The diagram includes only the papers that received NSF support at D=4 *in the global network* (which includes the EU). While all of these "startpoints" were located at D=4 in the Sankey diagram of the global network (Figure 5), some of these startpoints are now further from the patent boundary, at Distance values of 5 and 6. This demonstrates how the removal of the EU from the network makes some of the paths from grants to patents longer.

Figure 5 generates two core insights. The first is that less knowledge flows through the EU after excluding papers with a majority of authors. Some knowledge continues to flow through the EU because papers with a minority of EU authors remain in the network, which can be interpreted as an incompleteness of the shock, or the ability for teams with EU minorities to reorganize and continue their research. The second insight is that a larger quantity of knowledge flows through the U.K. and the rest of the world after excluding the EU. This dynamic indicates that the UK and other countries help to increase the resilience of the U.S.' knowledge supply to shocks affecting the EU.

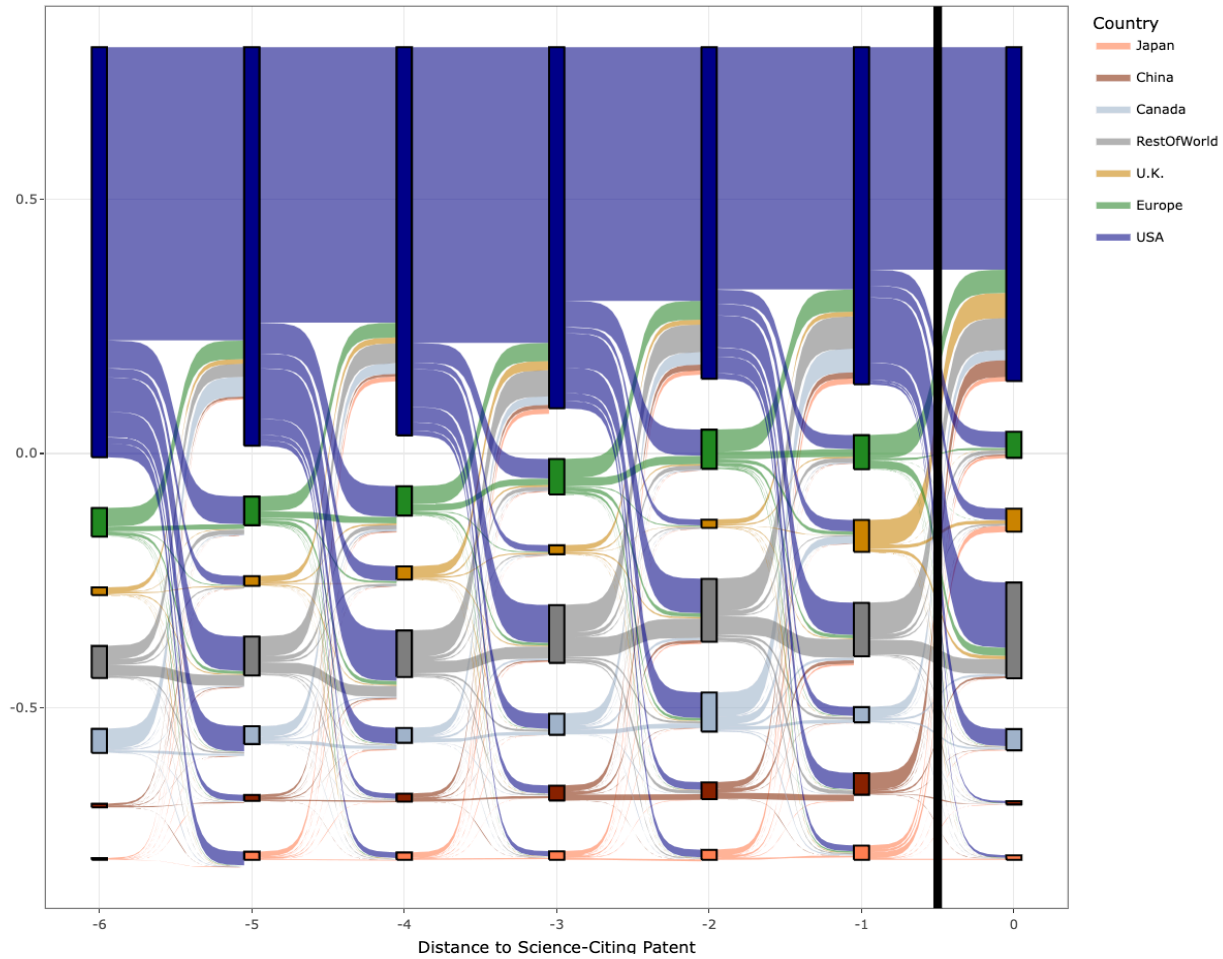
Figure 5: Knowledge Trajectories Supported by the NSF at D=4, Omitting Papers with Majority EU Authors



To demonstrate how the U.S. knowledge supply responds to larger shocks, Figure 6 presents a Sankey diagram of the shortest-path trajectories from science that was supported by the NSF at D=4 in the global network, after excluding all of the major science producers of Europe from the network (the EU plus Norway, Switzerland, and the U.K.). Again, there are two core insights produced by the figure. First, the U.S. no longer "offshores" the intermediate stages of knowledge development to other countries; the U.S. has a similar share of authors involved in knowledge production at all values of distance to the nearest patent. Second, other countries

provide a limited amount of substitution for the knowledge that had passed through the EU and the U.K.

Figure 6: Knowledge Trajectories Supported by the NSF at D=4, Omitting Papers with Majority Authors from Europe (EU27, GB, NO, CH)



4B) Quantifying the Resilience of the U.S. Knowledge Supply Chain

I compute two measures to quantify the resilience of the U.S. knowledge supply chain. The first is the share of the articles that are supported by NSF funding that are connected to patents in the shortest-path citation networks. This measure indicates whether inventors are able to convert NSF funding into patented technologies. I compute the connected share of supported articles for

the full global citation network and 6 alternative networks that exclude papers produced by different sets of authors: papers with majority Canadian scientists, papers with majority Chinese scientists, papers with majority EU27 scientists, papers with majority European scientists (EU27 plus GB, CH, and NO), papers with a majority of their authors located outside the U.S., and papers with any authors outside the U.S. The share of the NSF-supported scientists for each of these networks is presented in Table 1, as well as the difference (in percentage terms) between the share of connected NSF-supported networks in the full global network and in each of the counterfactual networks.

Table 1: Share of NSF-Supported Articles Connected to Patents (%) Under Conditions of Restricted Access to Global Science

	Network	Share Connected to Patents (%)	% Change in Connected Article Share
1	Full Global Network	32%	
2	No Canadian Science	31%	-3%
3	No Chinese Science	31%	-3%
4	No EU27 Science	27%	-16%
5	No European Science (Including GB, CH, and NO)	24%	-25%
6	No International Science	17%	-47%
7	No International Science (All authors must be in U.S.)	12%	-63%

Note: A paper is defined as being from a country or a region if a majority (≥50%) of its authors are from that region. The exception is the case of "No International Science (> 0 international authors)", where a paper is defined as international if 1 or more of its authors are outside the U.S. The change in grants realized is the change (%) in the share of grants that are realized under a specified condition relative to the no-restrictions condition.

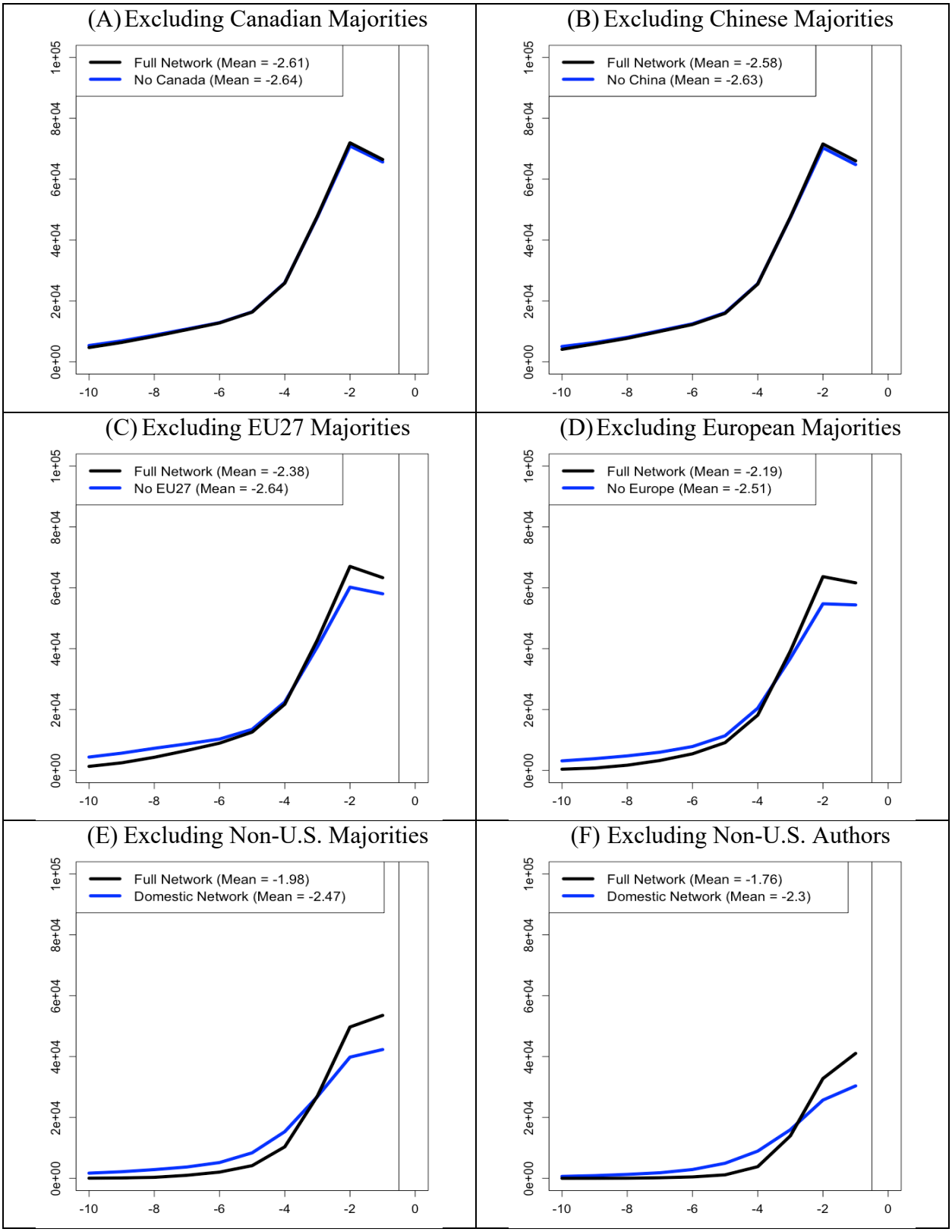
Table 1 shows that 32% of NSF-supported articles are indirectly or directly connected to patents in the global network. In the network that omits papers with majority Canadian authors, the share drops to 31%, a 3% decline. Omitting papers with majority Chinese authors also drops the share to 31%, while omitting majority-EU27 papers drops the share to 27% and omitting papers with majority European authors drops the share to 27%. When all international science is omitted, the share drops further, to 17% when majority-international authors are omitted, and to 12% when all papers with international authors are omitted. The latter results demonstrate that the U.S. citation graph has no domestic paths to connect 47% to 63% of its NSF-supported articles to patents that are connected to patents in the global network.

In addition to this measure of the share of NSF-supported articles connected to patents, I compute the change in the path length between NSF-supported articles and their downstream patents. This measure captures the potentially greater amount of required research effort required to convert NSF grants to patents in the absence of certain international research teams.

Figure 7 presents 6 panels, each of which contrasts the shortest path length from NSF-supported articles to their nearest patents using 6 hypothetical networks (shown in blue) to the global research network. Figure 7A contrasts the network that omits majority-Canadian networks to the global network, and shows that the path lengths are similar across the two networks; the mean only increases from 2.61 to 2.64 in the no-Canada network. Figure 7B presents a similar contrast for the network excluding Chinese-majority articles and generates a similar inference: removing China from the network does not substantially slow the conversion of NSF grants to patents. Figure 7C presents a related contrast omitting EU27-majority articles; here, the mean path length increases from 2.38 to 2.64, and the plot indicates that this increase is due to the lower number of paths of length 1, 2, and 3 in the no-EU network and the larger number of paths of

length 5 or more. Figure 7D, which contrasts a network excluding all European-majority papers (including GB, NO, and CH), generates similar, more pronounced effects. Finally, Figures 7E and 7F present contrasts with networks that omit international-majority papers and papers with one or more international author. In both of these contrasts, the path length between the NSF-supported papers and the nearest patents are substantially longer than they are in the global network.

Figure 7: Contrasts of Path Length from NSF-Supported to Nearest Patents for the Global Network and for Restricted Networks



5) Discussion

This paper constructed the global supply chain of science relied on by U.S. inventors to convert NSF-funded science into patents. The citations made by patents to scientific article were traced back along their shortest-length paths to NSF-supported articles, and the countries of the authors of the papers that fall along these paths were used to construct the U.S. knowledge supply chains. In counterfactual simulations, trajectories were traced from patents to NSF-supported paper while avoiding all papers produced by authors from countries of interest. In some cases, the trajectories were unaffected by the removal of such papers; in other cases, the trajectories were affected but were able to find alternative (and possibly longer) paths to the NSF-supported articles; in yet other cases, the trajectories were not able to reach their NSF-supported articles. Overall, the removal of international science causes 47% to 63% of the trajectories to irreparable break. For the remaining trajectories that can be reconfigured, the average path length increases by 25% to 31%.

This study has several limitations. First, it limits the analysis to the shortest length paths between papers and patents. This method has two core advantages, in that it reduces the complexity of the dataset to improve computational feasibility, and that it develops a coherent framework in which knowledge pipelines between documents are seen as passing between dyads. This dyadic approach also introduces an important limitation, because it does not describe the generative process of knowledge creation. For example, one might believe that all citations made by a focal paper, as opposed to just the citation that falls along the shortest path, were necessary for the creation of the focal paper. To the extent that actual knowledge creation is uncorrelated with the shortest-length paths in citation networks, the predictions generated by this method will lose accuracy.

Second, the current available data might understate the importance of Chinese science to U.S. innovation. U.S. patents increasingly cite applied Chinese (Wang, Li, and Wang 2024); however, more time must pass we will know whether U.S. patenting also relies of U.S. patenting on Chinese *basic* science (because decades must pass until basic science influences patenting). Currently, the methods used in this study do not attempt to deal with the temporal dynamics in knowledge flows between countries. One potential approach to this issue could be to study temporal trends in the citation probabilities between countries and to project those trends forward in time. For example, if U.S. patents are increasingly citing Chinese applied science, and if this increase is a linear function of time, we might expect that linear trend to continue. This solution, however, might not be as helpful in dealing with the potential growth of Chinese basic science.

Third, because of the right-truncation of knowledge trajectories, the current methods may overstate the number of NSF-patent trajectories that would be severed if the U.S. were isolated from global science. Because U.S. isolation would slow the speed of converting NSF-supported research into patents, some of the new paths from NSF-supported papers to patents may not have yet materialized by the conclusion of the patent dataset (2022). Therefore, it may be necessary to correct the estimates for right-truncation. Such a correction would likely cause the share of irreparably severed trajectories to decrease, but would also cause the mean increase in average path length to increase.

Finally, there are two clear directions in which to extend the research. First, the records of support from the NSF also include grant numbers, which can be linked to grant size (in dollars). With a few assumptions (such as fractionally assigning grant money to papers), it is relatively straightforward to determine the size of public investments in each article. Second, several studies have developed methods to determine the financial value of patents (Kogan et al. 2017; Hsu et al.

2021). Combining information on the financial value of public investments in basic science with information on the financial value of the downstream patents may allow the scale and the importance of international science to the U.S supply chain of science to be expressed in dollar terms.

References

Ahmadpoor, Mohamma, and Benjamin Jones. 2017. “The Dual Frontier: Patented Inventions and Prior Scientific Advance.” *Science* 357 (6351): 583–87.

Esposito, Christopher. 2023. “The Geography of Breakthrough Invention in the United States over the 20th Century.” *Research Policy* 52 (7).

Fleming, Lee, and Olav Sorenson. 2001. “Technology as a Complex Adaptive System: Evidence from Patent Data.” *Research Policy*. Vol. 30.

Hsu, David, Po-Hsuan Hsu, Tong Zhou, and Arvids Ziedonis. 2021. “Benchmarking US University Patent Value and Commercialization Efforts: A New Approach.” *Research Policy* 50 (1): 1040–76.

Jones, Benjamin F. 2009. “The Burden of Knowledge and the ‘Death of the Renaissance Man’: Is Innovation Getting Harder?” *Review of Economic Studies* 76 (1). https://doi.org/10.1111/j.1467-937X.2008.00531.x.

Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman. 2017. “Technological Innovation, Resource Allocation, and Growth.” *Quarterly Journal of Economics* 132 (2): 665–712.

Lin, Zihang, Lu Liu, and Dashun Wang. 2023. “SciSciNet: A Large-Scale Open Data Lake for the Science of Science Research.” *Scientific Data* 10 (215).

Marx, Matt, and Aaron Fuegi. 2020. “Reliance on Science: Worldwide Front-Page Patent Citations to Scientific Articles.” *2020* 41 (9): 1572–94.

Samila, Sampsa, and Olav Sorenson. 2010. “Venture Capital as a Catalyst to Commercialization.” *Research Policy* 39 (10): 1348–60.

———. 2011. “Venture Capital, Entrepreneurship, and Economic Growth.” *Review of Economics and Statistics* 93 (1): 338–49.

Wang, Junhan, Xibao Li, and Yangbo Wang. 2024. “The Contribution of Chinese Science to US Technological Advancement: Evidence from Patent Citation of Academic Papers.” *SSRN Working Paper #4802379*.

Wouden, Frank van der. 2020. “A History of Collaboration in US Invention: Changing Patterns of Co-Invention, Complexity, and Geography.” *Industrial and Corporate Change* 29 (3): 599–619.

Wu, Renli, Christopher Esposito, and James Evans. 2024. “China’s Rising Leadership in Global Science.” *arXiv: 2406.05917*.