Communication Costs in Science

Evidence from the National Science Foundation Network

Ezra G. Goldstein*

August 19, 2020

Abstract

How do communication costs affect the creation of scientific output? This study examines changes in scientific output and citation patterns following an institution's connection to the National Science Foundation Network (NSFnet), an early version of the Internet. Established in 1985 to connect five NSF-sponsored supercomputers, the NSFnet national internet backbone quickly expanded to universities across the United States by linking existing and newly-formed, wide-area regional computer networks. I estimate the effect of connection to the national internet backbone on citations per paper by exploiting plausibly exogenous variation in the connection times of the regional NSFnet networks. Following connection to the national NSFnet, average citations per paper increase by over ten percent relative to the pre-connection mean. Subgroup analyses reveal that the net effect was driven largely by middle- and top-tier institutions. Finally, I show that NSFnet connection led to a decline in interdisciplinary citations and an increase in within-field citations, but I find no evidence of increases in collaboration patterns.

Keywords: Communication Costs, NSFnet, Economics of Science, Interdisciplinarity

JEL Classification: O31, O33, I23, N72

---

*eggoldstein@fsu.edu. Department of Economics, Florida State University, 288 Bellamy Building, Tallahassee FL 32301, United States. I am grateful for helpful comments from E. Jason Baron, Shawn Kantor, Carl Kitchens, Luke Rodgers, and Douglas Van Houweling, as well as participants at the Florida State University Quantitative Economics Seminar and the Economic History Association 2019 Meetings. All errors are the sole responsibility of the author.

1 Introduction

Throughout history, invention and scientific progress have largely been the products of recombination, the utilization of existing techniques and technologies into novel ideas (Weitzman 1998; Mokyr 2002, 2005). It is then not surprising that technologies which reduce communication costs and improve access to existing knowledge, particularly the Internet, might play an important role in the creation of new knowledge. By using publication and citation patterns as a proxy, I examine how a decline in communication costs following connection to the early Internet affected the creation of new knowledge and flow of information among academic institutions. Specifically, I exploit plausibly exogenous variation in network availability induced by the introduction of the National Science Foundation Network (NSFnet).

Created in 1985, the NSFnet program linked together five NSF sponsored supercomputers across the United States. Shortly thereafter, the program was transformed into the first high-speed, wide-area, and general-purpose research and academic network (Frazer 1995). The NSFnet rapidly expanded to academic institutions across the United States by connecting both existing and newly-created regional computer networks to the national NSFnet backbone. Regional network connections occurred through several phases of the national NSFnet's various upgrades, with the major regional NSFnet networks coming online by 1988.

I rely on two sources of data to study the effect of early internet availability. First, I collect and digitize a detailed source on regional NSFnet networks that lists connected institutions and time of national NSFnet connection. Second, I complement NSFnet connection data with a database of publications made by the top U.S. institutions across twenty broad scientific disciplines from 1981–1999. I show that, prior to NSFnet connection, institutions connected early to the national NSFnet program followed similar trends in publication measures relative to later connected institutions.

In order to estimate the effect of a reduction in communication costs on academic publishing and citation patterns, I exploit plausibly exogenous variation of regional network connections to the NSFnet backbone. Baseline estimates indicate that connection to the national NSFnet increased average citations per paper by roughly 10 percent relative to the pre-connection mean. Additional estimates show connection to the national NSFnet primarily affected citations rather than altering the amount of published papers. Put differently, results suggest that reductions in communication costs did not alter the amount of knowledge produced, but potentially changed quality of articles and the flow of information.

Sub-group analyses reveal heterogeneity in the effect of NSFnet connection throughout various measures of the institutional quality distribution. Specifically, the results indicate that the observed increase in average citations per paper is driven by institutions at the middle and top of the quality distribution. Among both middle- and high-quality institutions, the net effect of connection to the national NSFnet on average citations per paper was similar. Conversely, I find that the increase in average citations per paper from connection to the NSFnet was not accrued to institutions in the bottom of the institutional quality distribution.

Finally, I explore plausible mechanisms that may account for the estimated increase in average citations per paper. Since collaboration between institutions often leads to higher quality scientific work, the documented increase in average citations per paper could be the result of changes in collaboration following NSFnet connection (Wuchty, Jones, and Uzzi 2007; Jones, Wuchty, and Uzzi 2008).[1] To test this, I examine whether connection to the national NSFnet influenced collaboration between pairs of institutions. I do not find significant evidence that connection to the national network led to an increase in collaboration.

Alternatively, increases in citations per paper could be driven by increased access to previously unknown research as reductions in communication costs may alter the flow of information among researchers. The ability to filter online information, combined with the researcher's constrained information capacity, may lead researchers to screen out interdisciplinary science in favor of discipline-specific, specialized knowledge (Van Alstyne and Brynjolfsson 1996; Rosenblat and Mobius 2004). In other words, as communication costs fall, preferences rather than geography or ease of access may determine scientific connectivity.

---

[1] Since 1960 the size of an average research team has expanded by 50 percent. An increasing amount of discovery is done by teams that frequently span the boundaries of an individual institution with the highest-impact research more often involving elite institutions (Wuchty, Jones, and Uzzi 2007; Jones, Wuchty, and Uzzi 2008). The dramatic rise of teams in science has been linked to the wide-spread adoption of internet technologies and scientific specialization (Adams et al. 2005; Jones 2009; Singh and Fleming 2010).

To study how early internet technology affected the flow of information, I examine patterns of citations between pairs of academic institutions both within and between broad scientific disciplines. I find that interdisciplinary citations declined on average while within-discipline citations grew following connection to the national NSFnet. The results provide evidence that reductions in communication costs seem to nurture specialization in science by enabling scientists to better access research within their own discipline.

This paper contributes to three related literatures. First, the paper adds to the literature examining the effects of information and communication technologies (ICTs) in science. Previous empirical studies restrict their analyses to narrow disciplines or publications appearing in discipline-specific, high-quality journals. While these studies are able to carefully trace collaboration histories and publication impact, they warn against generalizing their findings to other disciplines (Adams et al. 2005; Agrawal and Goldfarb 2008; Ding et al. 2010; Adams and Clemmons 2011). This paper expands the literature by studying the impact of information and communication technologies across 20 broad scientific disciplines and by shedding light on how a reduction in communication costs alters the interaction between disciplines. Consistent with theoretical predictions by Van Alstyne and Brynjolfsson (1996) and Rosenblat and Mobius (2004), the findings indicate that reductions in communication costs led researchers to further specialize within scientific disciplines and reduce interdisciplinary interactions.

The findings in this paper also contribute to the literature that studies the knowledge production function by underscoring the role of communication costs, the accessibility of information, and information technologies. Access to frontier knowledge influences the productivity of both academics and inventors alike (Iaria, Schwarz, and Waldinger 2018). Recent studies provide compelling evidence that reducing barriers to information increases scientific output. For example, Biasi and Moser (2020) show, for science books exposed to a weakening of copyright law, reductions in access costs significantly increase citations. Teodoridis (2018) shows that sudden reductions to access frontier knowledge facilitates collaboration and influences the composition of research teams.[2] This paper adds to the literature by doc-

---

[2] Additionally, Murray and Stern (2007) finds that, following the grant of intellectual property rights, papers associated with patents experience a decline in forward citations. Furman, Nagler, and Watzinger (2018) and Berkes and Nencka (2019) highlight the role of library systems that reduce costs to access

umenting a change in academic output following a sharp reduction in communication costs brought on by an early internet technology. Specifically, this paper highlights the importance of communication costs in determining information flows between scientific fields.

Finally, the paper is related to studies that examine the effects of information and communication technologies on productivity by documenting the effect of an early version of the modern Internet on academic output.[3] The findings differ from earlier studies in that they rely on variation in connection to a nationally-provided internet service rather than being inferred using self-reported data or time-period measures of ICT availability. Furthermore, much like other production processes explored in the literature, results presented in this paper suggest the effect of early ICTs in science were persistent and meaningful. This paper lends evidence to the claim that ICTs change production functions, not just for commodities and services, but also for knowledge.

2 Background

Prior to the NSFnet, computer networking at universities largely consisted of campus networks: linked computers and mainframes on the campus of an individual university. At that time, computer-networking technology was in its infancy and communication across machines, programs, or protocols was a technological challenge.[4] The disconnect between adopted technologies made large-scale, wide-area networking nearly impossible. As a result, few universities were connected to one another beyond a handful of regional computer networks that rarely crossed state boundaries.[5] With the NSFnet project, the NSF set out to

---

previously unknown work and provide access to more technologically diverse resources. Gross (2019) shows reduced access to patents from the USPTO's secrecy program during the Second World War diminished follow on innovation.

[3] Prior studies have shown the effectiveness of ICTs in re-configuring the workplace by altering management practices, reducing information frictions and search costs, facilitating better employment matches, and the utilization of idle inputs (Bloom, Sadun, and Van Reenen 2012; Forman and Van Zeebroeck 2012; Bloom et al. 2014; Kuhn and Mansour 2014; Horton and Zeckhauser 2016). In addition, the productive benefits from the advent of ICTs largely accrue to high-skilled labor (Akerman, Gaarder, and Mogstad 2015). For a comprehensive overview on the study of ICTs in the economy and "digital economics" see Goldfarb and Tucker (2019).

[4] The breakthrough came in 1974 when Robert E. Kahn and Vinton Cerf created Transmission Control Protocol / Internet Protocol (TCP/IP) for use on the Department of Defense's ARPANET.

[5] Prior to the NSFnet, regional networks existed primarily to facilitate sharing of computational resources between state-wide or local institutions (such as Michigan's MichNet and Los Angeles' Los Nettos). Early

create a "network of networks" by enabling campus computer networks to communicate with each other across the new, national "inter-net" (Frazer 1995).

The NSFnet began in earnest in 1985 when the NSF began funding the construction of five super-computing research centers. Shown in Figure 1, the first phase of the NSFnet backbone connected the five computing centers and and the National Center for Atmospheric Research (NCAR). Each super-computing center (and later several universities) hosted backbone nodes in the national NSFnet infrastructure. Backbone nodes functioned as a hub for both pre-existing or newly-created regional NSFnet computer networks which connected individual campuses to the early internet. Figure 2 shows the 1986 NSFnet backbone that facilitated connections between individual campuses by linking together regional NSFnet computer networks tethered to national NSFnet backbone nodes.

The national network was constructed by employing networking technologies developed by the ARPANET and a novel three-tiered architecture, shown in Figure 3, which linked individual institutions to the NSFnet via regional network connection.[6] Importantly, the three-tiered architecture allowed the national NSFnet program to focus on providing frontier network infrastructure to users while empowering regional networks to provide connectivity for individual institutions and handle day-to-day operations (Frazer 1995). Unlike other wide-area networks of the day that were connected at the request of individual institutions or researchers, national NSFnet connections were managed by a board of computing directors from regional networks and connected institutions with little influence coming directly from researchers. Ultimately, the NSFnet's innovative design allowed newly linked machines on campus networks to communicate via the national backbone. The result was a remarkably reliable, high-speed network able to transmit information across the country.

---

adopters of computer networking had begun sharing comments on concurrent research projects as early as 1964, however communication was constrained by proximity to local mainframes (Hafner and Lyon 1998, p. 13).

[6] The NSFnet and much of the modern Internet has its roots in the ARPANET, a continental defense research oriented computer network established by the Department of Defense's Advanced Research Projects Agency (ARPA) in 1969. Originally designed to develop and test new computer networking infrastructure, the ARPANET quickly became an important tool for communication among ARPA affiliated researchers. Many technologies developed by the ARPANET project, such as e-mail, file transfer protocol (FTP), and transmission control protocol/internet protocol (TCP/IP), are still employed today. The network, however, was a restricted access network available only to a few universities working closely with ARPA projects (Frazer 1995).

Although the NSF itself was specifically formed to cater to science and engineering, the NSFnet program was created to be a general-purpose research and education computer network. Reflected by the network's acceptable-use policy, the national network placed few restrictions on scholars and welcomed researchers of every stripe.[7] As a result, early users of the NSFnet came from a wide range of scientific disciplines and benefited from innovative communication services made available on the national network.

While the NSFnet first meant to grant geographically dispersed researchers access to supercomputing technology, the most utilized technologies were file transfer and electronic mail services (Claffy, Polyzos, and Braun 1993). Prior to the wide-scale availability of information and communication technologies, researchers did not immediately share results of projects with colleagues outside of their own departments. Collaborators engaged in costly, sluggish information sharing; arranging to meet at upcoming conferences, sending data by mail, and accumulating the latest working papers via subscription. Services provided by the NSFnet enabled researchers, many for the first time, to share files, publish digital working papers, send articles, and work in real-time with distant colleagues (Frazer 1995).

Adoption of the new networking service among academics was so popular and unexpected that just a year after its launch the NSFnet expanded. In 1987 the NSF selected the Michigan Education and Research Information Triad (MERIT), a Michigan non-profit regional computer networking consortium, along with IBM and MCI to manage the NSFnet and its expansion.[8] Under MERIT's management, the ultimate goal was simple: "generalized connectivity" (Frazer 1995). Expansion of the NSFnet manifested through the establishment of new regional computer networks, connection to pre-existing regional networks, and an upgrade of the national NSFnet backbone infrastructure to high-speed T-1 lines.[9] Typically, regional networks began as collections of universities and expanded as regional network

---

[7] The acceptable use policy read "NSFNET Backbone services are provided to support open research and education in and among US research and instructional institutions, plus research arms of for-profit firms when engaged in open scholarly communication and research. Use for other purposes is not acceptable." *Source:* `http://www.cybertelecom.org/notes/nsfnet.htm#aup`

[8] MERIT had been selected from a pool of five proposals solicited by the NSF. MERIT stood in a unique position to win the NSFnet project. The Michigan-based education consortium had been managing the longest running regional computer network, MichNet, since 1971.

[9] The original NSFnet operated on 56 kilobit lines while the T-1 lines operated on 1.5 megabits. This upgrade effectively enhanced the speed of the network thirtyfold.

managers saw fit, with MERIT providing managerial and technical support.

By mid-1988 over 170 research and education institutions were, many for the first time, connected to a high-speed, high-performance communications service. One month following the completion of the T-1 network, traffic doubled and increased by ten percent each month thereafter (Frazer 1995). Much of the traffic on the network was concentrated between pairs of individual institutions, with 0.28 percent of campus pairs generating 46.9 percent of network traffic (Claffy, Polyzos, and Braun 1993).[10] Figure 4 shows the completed T-1 NSFnet backbone. In total, the NSFnet hosted sixteen regional networks along with several sub-regional networks that connected over 50,000 individual computer networks sometime between MERIT's award and the eventual transition of the NSFnet to the modern Internet in 1995.

3 Data

In order to examine the effect of national NSFnet connection on knowledge production, I collect details of regional NSFnet network connection to the national network and publication data from the top academic institutions in the United States.

3.1 NSFnet Connection Data

To collect NSFnet connection timing, I rely details of regional NSFnet networks listed in *The User's Directory of Computer Networks* (LaQuey 1990). Designed as a "road atlas", *The User's Directory of Computer Networks* details academic- and research-computer-network capacity of the United States in the late 1980s. Crucially, the directory lists the member institutions and hosts of various computer networks. The section of the directory dedicated to the NSFnet details a brief history and organization of the national NSFnet project including each regional network connected to the national NSFnet backbone. The directory provides several characteristics of each regional NSFnet network such as institutions that constitute

---

[10] Traffic generated between institutions was not symmetric. As Claffy, Polyzos, and Braun (1993) highlight, in May of 1992 "the most heavily used link for the month, College Park to Houston, had utilization almost always exceeding 20% (for fifteen-minute intervals) and more than 50% during peak hours of the day. Interestingly, the reverse direction, Houston to College Park, had almost uniformly lower utilization."

the network, year of connection to the national NSFnet backbone, and maps of network topology. When available, I cross-reference each map with membership lists to assure a complete listing of members to the regional NSFnet networks.[11]

Figure 5 highlights the directory entry for MIDnet, a regional NSFnet network connected to the national NSFnet. MIDnet provides an excellent example of both pre-existing and newly-created regional-networking. MIDnet itself was founded in the spring of 1985 at the start of the NSFnet program, but connected to the national NSFnet backbone in September of 1987. Also shown in Figure 5, MIDnet was also comprised of sub-regional networks that serviced institutions from individual states. The University of Missouri Network, for example, was a sub-regional computer network comprised of institutions within the Missouri university system. Prior to the NSFnet, users within state university research networks had little resources to access computer networks apart from their own. With connection to the national backbone, regional NSFnet networks and sub-regional networks like the University of Missouri Network became a part of the new, national NSFnet.

An important caveat is that the connection years reported in *The User's Directory of Computer Networks* may not represent the true connection to the NSFnet. For instance, several regional networks report the year the network is "fully operational," but it is not clear that previous stages of the network would not provide some of the same tools given by a fully operational network. In addition, I rely on regional network membership at the time of the directory's publication. In other words, an institution may have joined the regional network following the establishment of the regional network or following connection to the national NSFnet. Since I am unable to observe the original members of all regional networks connected to the national NSFnet, I assign to each institution the year of earliest regional NSFnet connection.[12]

---

[11] In the event that a connection year is not listed for a regional network I rely on information published on *LivingInternet*, a web-publication dedicated to the history of the internet, to assign connection years. See `https://www.livinginternet.com/i/ii_nsfnet.htm`

[12] The University of Arizona, for example, is listed as having connected to both WESTnet in the southwestern United States and the John von Neumann Center (JVNC) in Princeton via satellite connection. WESTnet connected to the national NSFnet in 1988, while JVNC connected to the national network in 1986 as an originally funded supercomputing center. In this case, the University of Arizona is assigned 1986, since it is earliest national NSFnet connection.

3.2 University Publication Data

Next, I pair the timing of NSFnet connections with publication data from the *NBER-Rensselaer Polytechnic Institute Scientific Papers Database* assembled by Adams and Clemmons (2008) (henceforth referred to as AC). The AC database contains aggregated publication statistics from more than 2.5 million scientific papers by the top 110 American universities during 1981–1999 and contains three key data sources that enable me to study the effect of the NSFnet on scientific output.

First, the AC publications panel contains data on institution-field observations that aggregate each paper by publication year, by its affiliated institution, and by one of twenty CASPAR20 field codes determined by the journal in which the paper appears.[13] To avoid multiple counting of scientific output the AC publications panel "fractionalizes" published papers and forward citations. Fractionalized measures assign an equal share of the publication to each institution associated with the publication. Borrowing an example from Adams and Clemmons (2008), if researchers at Harvard write a paper by themselves, Harvard receives the full paper and all forward citations. If researchers at Harvard collaborate with researchers at Yale, both institutions receive half of the paper and forward citations. In the analysis that follows, I rely on the fractionalized variables as a measure of scientific output within a given institution-field observation.[14]

Second, the AC citations dataset reports pairwise directional citations between institution-field observations throughout the sample period. For example, in 1983 Stanford papers published in biology cited Harvard papers published in medicine 19 times; meanwhile, in the same year, Harvard papers published in medicine cited Stanford biology papers 32 times. The citations dataset enables me to examine how NSFnet connection affected citation patterns, both within and across scientific disciplines, between pairs of concurrently connected institutions.

---

[13] The AC database classifies each paper by 88 scientific fields which I aggregate to the CASPAR20-level using a crosswalk provided by Adams and Clemmons (2008). Classifying each discipline to the National Science Foundation's CASPAR field codes makes it possible to combine the AC database with NSF survey data described in Section 3.3. The twenty CASPAR fields are listed in Table A.1.

[14] Adams and Clemmons (2008) show that whole counts of publications and citations overstate the total amount of national scientific output by approximately 18 percent. I show below in Section 4.2 that the baseline results presented in this paper are robust to using whole counts of total citations, published papers, or citations per paper.

Third, the AC collaboration dataset details the total number of collaborations between institutions within a scientific discipline. For example, in 1983, Stanford and Harvard collaborated on five papers published in medicine. Since the AC data codify scientific discipline by journal, there can be no comparison of collaboration across fields. The collaboration dataset enables me to examine how NSFnet connection affected collaboration rates between pairs of concurrently connected institutions.

3.3 Supplemental Data

I complement the publication information and NSFnet connection timing with detailed, time-varying data covering research inputs by each institution. I collect counts of awarded doctorates from the National Science Foundation Survey of Earned Doctorates during the sample period. In addition, I collect the number of graduate students and postdoctoral scholars from the National Science Foundation - National Institutes of Health Survey of Graduate Students and Postdoctoral Scholars in Science and Engineering. Number of graduate students, postdoctoral scholars, and awarded doctorates are collected at the institution-field-year level.

Next, I collect data on R&D expenditures from the National Science Foundation Survey of Higher Education Research and Development Expenditures. The survey provides expenditure data by each institution at the CASPAR12 level throughout the sample period.[15] For each institution, I collect total R&D expenditures and R&D expenditures funded by federal sources.

Additionally, I collect the total number of employed faculty and total salary outlays by academic rank. Employed faculty and salary outlays are available at the institution-level throughout the sample period. Data on faculty employment and salary outlays are collected from the National Center for Education Statistics.All monetary values reported are in millions of real 2015 dollars.

---

[15] The CASPAR12 field codes are slightly more general than the CASPAR20. The primary difference stems from the aggregation of engineering and earth sciences. For a brief discussion of the CASPAR field code hierarchy see Section A.1.

3.4 Final Sample

Of the top 110 U.S. universities in the AC database, 102 are contained within some regional computer network listed in *The User's Directory of Computer Networks*. A further six institutions could be located through online records. The remaining two institutions that could not be located in the directory or via an online source, but were likely connected to the regional NSFnet networks, are excluded from the analysis.[16]

The final three datasets employed in the analyses contain balanced panels that cover the 108 top institutions across twenty broad scientific disciplines from 1981–1999.[17] In Section 4, I rely on the AC publications panel from 1981–1999 linked to the timing of national NSFnet connection as well as supplementary data to proxy for research inputs. Altogether, the AC publications panel consists of 2,130 individual institution-field observations from 1981–1999.[18]

Sections 5.1 and 5.2 utilize the AC collaboration and citation datasets, respectively. The AC collaboration data consist of 44,026 institution pairs and the AC citation data consist of 589,720 institution-field pairs through the sample period. I link national NSFnet connection times to each institution within their institution pairs and assign to each pair the earliest year of concurrent NSFnet connection. Figure 6 presents the geography of NSFnet connections. Among the connected 108 institutions, 45 were connected to a regional network that joined national NSFnet in 1986, 44 in 1987, ten in 1988, six in 1989, two in 1990, and one in 1991.

---

[16] Online records typically are rich-text archives of university bulletins. I search for the earliest mention of either the NSFnet or a regional network and assign the year of bulletin publication as the year of NSFnet connection. I am unable to find the NSFnet connection timing for The University of Connecticut which likely connected to the New England Area Research Network (NEARNet) or the New York area network (NYSERNET) sometime after the publication of the directory. In addition, I am unable to identify the NSFnet connection timing for the Oregon Health Science University.

[17] Not all institutions are represented across each scientific discipline. If an institution had no publications in a given scientific discipline across the nineteen years covered in the AC data, they are not represented. For example, the University of Texas San Antonio Health and Science Center did not publish in the field of chemical engineering throughout the sample period, so the institution-field pair is not represented in the data. While this may initially raise the concern of potential selection bias, the institution-field fixed effect differences out the observations where no publication is made throughout the sample period.

[18] For specifications that rely on time-varying controls, one institution is omitted for lack of available data. Woods Hole Oceanographic Institution is a special case in the AC data that has no NSF CASPAR survey data available. The Woods Hole Oceanographic Institute case accounts for the drop in observation count when including control variables.

3.5 Summary Statistics

Figures 7, 8, and 9 present annual trends in observable variables by connection group. The *First Connected* group contains the 45 institutions that had been connected in 1986 and the *Later Connected* group contains all other institutions. At a glance, it appears that first connected institutions receive more total citations, publish more papers, receive more citations per paper, have greater budgets for faculty and R&D, enroll more graduate students, employ more postdoctoral scholars, and award more doctorates than their later connected counterparts. The figures provide evidence that institutions connected earlier to the NSFnet were not on dissimilar trends of observable characteristics compared to their later connected peers.

Table 1 presents the pre-NSFnet mean and standard deviation (in square brackets) of several observables. Variables are grouped by their level of aggregation with CASPAR20, CASPAR12, and institution-level variables presented in Panels A, B, and C, respectively. Column 1 displays the pre-NSFnet program mean across all (eventually) connected institutions prior to the beginning of the national NSFnet program in 1986. Columns 2 and 3 present the pre-NSFnet means separated by connection group. Means presented in the table confirm the descriptive differences between connection groups. This is expected given that many large research institutions were the leaders of the national networking effort and since the NSFnet was focused on supporting academic research, major institutions were connected via the NSF sponsored supercomputing networks.

One may be concerned that the timing of connection to the NSFnet is correlated with pre-existing conditions that may influence the future path of an institution's scientific output. Put differently, universities that connected early to the burgeoning internet may have been on different trends relative to universities that connected later and estimates of a national NSFnet effect may capture the difference. To investigate this further, I estimate the following equation:

$$Y_{iftg} = \Theta FirstConnected_g + \tau_t + \varepsilon_{iftg} \tag{1}$$

where Y_{iftg} are dependent variables and covariates listed in Table 1 for a university i in field f that is a part of connection group g during year t. Column 4 of Table 1 displays the estimates of Θ (standard errors in parentheses) two years before connection (i.e. $t = -2$). Examining the difference in means two years prior to connection avoids a concern that first connected institutions may have been affected by the NSF's supercomputer program that transformed into the national NSFnet project. The estimates largely confirm the initial observation that institutions first connected to the NSFnet did indeed produce higher-quality work (as measured by citations per paper), spent more on faculty and R&D, employed more postdocs and faculty, and awarded more doctorates on average.

While early connected institutions, on average, differ across observables from later connected institutions, the two groups may still follow similar trends before their connection to the national NSFnet. Column 5 of Table 1 displays the results of estimating Equation 1 on the differences between two and three years prior to connection. The estimates indicate that, while the differences in means between connection groups are pronounced prior to the national NSFnet, there is little evidence of differences in the trends of observables leading up to connection to the national NSFnet.

4 Empirical Strategy and Results

4.1 Baseline Estimates

To examine the effect of the NSFnet on scientific output I employ the AC publications panel dataset and estimate the following equation:

$$Output_{ift} = \beta NSFnet_{it} + \mathbf{X}'_{ift}\boldsymbol{\alpha} + \gamma_i + \nu_f + \tau_t + \varepsilon_{ift} \tag{2}$$

where i denotes the institution, f the scientific field, and t the year. The dependent variable $Output_{ift}$ denotes annual measures of an institution's academic output by scientific field. Academic output is measured annually at the institution-field level by average citations per paper, fractional citations, and fractional papers. $NSFnet_{it}$ is the independent variable of interest that takes the value of one for institution i in the year following national NSFnet connection. The specification also includes institution (γ_i) , scientific field (ν_f), and year (τ_t) fixed effects. Institution fixed effects account for any institution-specific time-invariant factors that may affect research output, such as an institution's preferences for research in a given subject or propensity to adopt frontier technologies. Scientific field fixed effects capture any field-specific characteristics that are time invariant. Year fixed effects control for any time-varying shocks that are common across all institutions and disciplines. Finally, $\mathbf{X}_{ift}$ is a vector of annual institution-field-level covariates to proxy for research inputs. Standard errors are clustered at the institution level to allow for correlation between between scientific disciplines within the same institution across time.

The coefficient of interest in Equation 2 is β which represents the average change in academic output, such as average citations per paper, following connection to the NSFnet. The average change in academic output is identified by exploiting the staggered timing of national NSFnet connection. To identify the effect of NSFnet connection on academic output, two assumptions must be satisfied. First, conditional on covariates and fixed effects, the timing of NSFnet connection is exogenous. Second, in the absence of NSFnet connection, academic output at first-connected institutions would have evolved in a similar manner to later-connected institutions before their eventual connection. As discussed in Section 2, individual institutions had little power over their connection to the national NSFnet backbone and relied on connection to the new national research-network via their regional computer network. The nature of regional network connections to the national NSFnet lends some evidence that the timing of national NSFnet connection is plausibly exogenous to the researchers at connected institutions. Moreover, Section 3.5 provides evidence that observable characteristics of first-connected and later-connected institutions were evolving similarly prior to the introduction of the NSFnet.

Table 2 presents estimates of β for three measures of academic output with institution-clustered standard errors shown in parentheses. Column 1 of Table 2 presents the result of estimating Equation 2. The baseline estimate indicates that, following connection to the NSFnet, institutions publishing in a given field saw an average increase of 0.55 citations per paper. Relative to the pre-NSFnet connection mean of 5.65 citations per paper, this corresponds to roughly a ten percent increase in average citations per paper as a result of national NSFnet connection. Column 2 of Table 2 presents estimates nearly identical to the baseline after controlling for time-varying factors within scientific fields that are constant across all institutions.

Although institutional evidence presented in Section 2 suggests the establishment of the NSFnet was not driven by output at individual institutions, one may be concerned that the NSFnet coefficient may be correlated with some variables omitted from the estimation. For instance, the estimate of national NSFnet connection could be biased if institutions' selections of inputs to research is correlated with the timing of regional NSFnet connection to the national backbone. To assuage concerns that the effect of connection to the national NSFnet simply reflects omitted inputs to the research process, I include several time-varying measures of research inputs. Notably, the estimate presented in Column 3 of Table 2 is larger in magnitude, yet not significantly different from estimates presented in previous specifications.

Finally, given that exploiting the variation of NSFnet connection amounts to variation in the university over time, one may be concerned that the baseline estimates are simply a reflection of a general trend in research output or some unobserved difference in trends between institutions. To address this, I include an institution-specific linear trend. One appealing feature of the inclusion of the linear trend is that it allows the relaxation of the assumption of similar trends. While restricted to be linear, the specification allows institutions to be on different trends. Column 4 of Table 2 shows shows the inclusion of institution-specific linear time trends. The estimated coefficient of NSFnet connection falls to a four percent increase and remains statistically significant at conventional levels. In other words, a general time trend cannot fully explain the estimated effect of NSFnet connection.

Additional estimates in Table 2 suggest that the estimated effect of national NSFnet connection on average citations per paper are driven by changes in citations rather than total papers. Estimates of Equation 2 with total citations as the dependent variable reveal an average increase in total citations between ten and 30 percent. Conversely, the estimated effect of national NSFnet connection on total papers is small, negative, and, when accounting for research inputs, insignificant. Taken together, the estimates presented in Table 2 suggest the national NSFnet increased average citations per paper and this change was driven by a relative increase in citations.

4.2 Robustness Checks

An additional concern is that using concurrent measures of research inputs may fail to capture the lag in the publication process faced by academic institutions. Estimates of national NSFnet connection on scientific output would be biased if institution-level, pre-NSFnet research inputs played a role in both determining regional network connections to the national backbone and academic output. To address this concern, I include lags of research inputs in Column 1 of Table 3. Reassuringly, the estimates are largely unchanged.

Next, instead of fractionalized measures, I estimate the effect of national NSFnet connection on whole counts of citations per paper, citations, and papers in Column 2 of Table 3. Recall that the fractional measures of publications divides the total amount of publications or citations equally to all institutional collaborators. The fractional count itself could be affected by a reduction in communication costs from the national NSFnet if the national network influenced the propensity to collaborate. Resulting estimates, using whole-count measure of academic output, presented in Column 2 are not significantly different from the baseline estimates.

One may be concerned that the documented NSFnet connection effect reflects the availability of earlier computer networking technologies. To address this concern, I include controls for another early wide-area computer network, BITNET, studied by Agrawal and Goldfarb (2008) and Ding et al. (2010). BITNET was an alternative computer network aimed to facilitate communication among universities. Unlike the NSFnet that managed connections by linking regional computer networks, individuals and universities connected to BITNET by requesting and leasing telephone lines (Gurbaxani 1990).

To account for any effects that may be due to BITNET connection, I match institutions within the AC database to their earliest BITNET connection year.[19] Column 3 of Table 3 includes controls for variation in BITNET connection timing. Notably, the estimates remain similar to those shown in Table 2 which suggests availability of other early computer networking technologies cannot explain the documented NSFnet connection effect.

---

[19] Of the 108 institutions, 98 appear in the BITNET connection data made available by Agrawal and Goldfarb (2008). For institutions that do not appear in the BITNET connection data, I code BITNET connection year as missing and omit them from the analysis.

4.3 Placebo Test

A remaining concern is that the observed national NSFnet connection effect is an artifact of assigned connection time. In other words, one might worry that the estimates presented above would be no different from a random draw of connection years. Furthermore, as highlighted by Bertrand, Duflo, and Mullainathan (2004), serial correlation can bias standard errors in typical difference-in-difference style analyses.

To address these concerns, I borrow from Chetty, Looney, and Kroft (2009) and conduct a nonparametric permutation test of the effect of connection to the national NSFnet.[20] For each regional network connected to the national NSFnet, I randomly assign a placebo national NSFnet connection year during the sample period and re-estimate the specification presented in Column 1 of Table 2. I conduct 500 repetitions of this exercise in order to generate an empirical distribution of placebo NSFnet connection.

Figure 10 shows the empirical distribution of the placebo estimates as well as the actual estimated effect of national NSFnet connection. The figure highlights that the actual estimated effect is far larger in absolute value than the placebo estimates. Moreover, as shown in Figure A.2, the actual estimated effect remains in the top of the empirical distribution regardless of specification.

4.4 Event-Study

The standard difference-in-differences style approach that assumes a sudden and constant response to a policy variable may be misspecified if the dynamics of the policy deviate from a simple binary shift. To explore the potential dynamic effects of connection to the NSFnet, I estimate the following equation:

$$CitsPerPaper_{ift} = \sum_{j \neq -1} \beta_j \mathbb{1}\left\{t - NSFnetYear_i = j\right\} + \mathbf{X}'_{ift}\boldsymbol{\alpha} + \gamma_i \times t + \nu_f \times \tau_t + \tau_t + \varepsilon_{ift} \quad (3)$$

---

[20] The permutation test closely follows that performed the study of Internet arrival in Africa by Hjort and Poulsen (2019). Hjort and Poulsen (2019) largely borrow the methods developed by Chetty, Looney, and Kroft (2009). While Chetty, Looney, and Kroft (2009) construct an empirical distribution from all possible permutations, Hjort and Poulsen (2019) assign 500 randomly chosen arrival dates for treatment.

where the indicator function takes the value of one when a university is j years from their NSFnet connection year. Included in this specification are six leads and eleven lags, therefore the coefficient of β_j represents the coefficient of the jth lead or lag.[21] Unlike Equation 2 which imposes a constant and immediate effect of connection (i.e. $\beta_j = \beta, \forall j > 0$), this specification uncovers the dynamic effect of connection by allowing for the estimation of the individual $\beta_j's$. Equation 3 extends the specification presented in Column 5 of Table 2 to an event-study framework. The omitted period is $(t = -1)$ so that all estimated coefficients are relative to the year before national NSFnet connection and identified by the staggered regional NSFnet connection timing described in Section 2.

Figure 11 plots the estimated $\beta_j's$ from equation 3 as well as the 95% confidence intervals. The estimated effect of connection on citations per paper manifests roughly two to three years following connection and continues through time. The individual dynamic effects are notably larger than those found in Table 2. Simply averaging the coefficients following connection to the NSFnet yields an increase of roughly 1.5 citations per paper. Relative to the average 5.65 citations per paper prior to the NSFnet program, the estimates suggest a 25 percent increase in citations per paper following connection to the NSFnet.

4.5 Heterogeneity by Institutional Quality

The impact of reductions in communication costs on citations per paper may vary by institutional quality. Reductions in communication costs in science may enable researchers from lower-quality institutions to better reach their peers at higher-quality institutions, and may enable research at the top to expand to a broader audience. To explore any heterogeneous effect, I generate quartiles of total salary outlays per faculty before the NSFnet program.[22]

I modify Equation 3 to include an interaction between the event-study coefficients and

---

[21] Where the sixth lead is equal to one for all leads greater than six and the eleventh lag is equal to one for all lags greater than eleven. The cut-offs are determined by the first leads and lags where there is a balanced panel of institution-field observations. This amounts to a "binned" event-study discussed in Schmidheiny and Siegloch (2019)

[22] The estimates to not depend on measures of quality. Estimates are similar if instead I generate quartiles based on pre-NSFnet average citations per paper, pre-NSFnet R&D expenditures, or pre-NSFnet total citations.

dummy variables that enumerate the quartile of average faculty salary. Figures 12a - 12d plot the total connection effects over time for each quartile. I find that the net effect of connection to the NSFnet is similar between the 2nd, 3rd, and highest quartiles of institution average faculty salary. However, I find no statistically significant effect of connection to the NSFnet on the lowest quartile of pre-NSFnet average faculty salary.

Thus far, the findings suggest that the effect of the NSFnet connections bolstered the citations for top-tier institutions. Earlier studies show, within narrow scientific disciplines, that communication technologies primarily benefit middle-tier academic institutions by fostering links with the elite (Agrawal and Goldfarb 2008; Ding et al. 2010). These findings often rely on samples of institutions throughout the distribution of institutional quality. The data employed in my analysis differ in that they capture only the highest quality institutions - the top 110 publishing institutions from 1981–1999. Compared to Agrawal and Goldfarb (2008), the institutions studied in this paper represent roughly the top half of the institutional quality distribution.

5 Collaboration and Citation Patterns

Taken together, the net effect of the NSFnet suggests that reductions in communication costs had a significant, lasting effect on academic output from top-tier institutions. The widespread, national internet backbone established by the NSFnet appears to have caused a fundamental shift in average citations per paper. Further, the change in average citations per paper seems to stem from a change in citations and not in the number of published papers.

However, the mechanisms by which the reductions in communication costs affect citations are ambiguous. On the one hand, reductions in communication costs may enable scientists across broad disciplines to better access each other's work, increasing inter-disciplinary science and thereby creating a "scientific global village" — a virtual academic community free from geographic boundaries (Van Alstyne and Brynjolfsson 2005). Put differently, by making information widely accessible, the NSFnet may have bolstered inter-disciplinary science by allowing researchers to more easily search across unfamiliar scientific fields.

On the other hand, reductions in communication costs may cause scientists to further specialize within their field and reduce interdisciplinary science. Theoretical predictions put forth by Van Alstyne and Brynjolfsson (1996) and Rosenblat and Mobius (2004) suggest that improvements in digital connectivity, the ability to easily filter information, and the researcher's constrained information capacity can lead to scientific specialization. Faced with boundless digital resources, the researcher may screen out interdisciplinary information in favor of discipline-specific, specialized knowledge. In other words, as information technologies reduce communication costs, preferences rather than geography or technology may determine scientific connectivity (Van Alstyne and Brynjolfsson 2005).

At the same time, reductions in communication costs may decrease the costs to collaborate across institutions. Given that teams increasingly produce more highly-cited research than individual researchers, the observed effect of national NSFnet connection may be, in part, driven by changes in collaboration across institutions (Wuchty, Jones, and Uzzi 2007; Jones, Wuchty, and Uzzi 2008). In the remainder of the paper I explore these plausible mechanisms for observed effect of the national NSFnet.

5.1 Collaboration Pairs

The documented increase in average citations per paper could be the result of an increase in collaborations between institutions as a result of reductions in communication costs. Internet technologies have been shown to foster collaboration and augment productivity among the highly-skilled (Agrawal and Goldfarb 2008; Forman and Van Zeebroeck 2012; Bloom, Sadun, and Van Reenen 2012; Bloom et al. 2014; Akerman, Gaarder, and Mogstad 2015). Increases in collaboration could lead to higher quality work and garner a larger number of citations (Wuchty, Jones, and Uzzi 2007). Furthermore, high-impact science is more often the result of ever-growing research teams that more often span the boundary of individual institutions (Jones, Wuchty, and Uzzi 2008). If the NSFnet significantly reduced the costs to collaborate, then the observed change in average citations per paper may be due to changes in collaborations between institutions.

I examine whether connection to the NSFnet between pairs of institutions altered the likelihood of collaboration by employing the AC collaboration dataset.[23] Specifically, I estimate the following equation:

$$Collaboration_{dft} = \Omega NSFnetBoth_{dt} + \eta_d + \theta_f + \varepsilon_{dft} \qquad (4)$$

where d denotes the pair of collaborating institutions i and j. The dependent variable, $Collaboration_{dft}$, takes the value of one if there is any collaboration between institutions i and j of pair d within scientific field f during year t. Here, *NSFnetBoth* takes the value of one when both institutions i and j of pair d are connected to the NSFnet in year t. Institution-pair fixed effects (η_d) account for any time constant factors that may influence collaboration between institution pairs. Discipline fixed effects (θ_f) account for time-constant differences across scientific disciplines that may influence collaboration. Standard errors are clustered at the institution-pair level to allow for correlation between collaborating institutions across time.

Equation 4 tests if connection to the NSFnet between pairs of institutions influenced the likelihood of collaboration. Table 4 displays the results of estimating Equation 4 by a linear probability model. C collaborations between institution-pairs is a count variable with a mass at zero. Of the 836,494 institution-pairs within disciplines 76 percent have zero collaborations and of the non-zero collaborations, the median number of collaboration is one.[24] I estimate Equation 4 where the dependant variable is an indicator if there are *any* collaborations between institutions within an individual field.[25]

The baseline estimates of Ω presented in Column 1 indicate that the probability of collaboration between institution-pairs declines by 0.3 percent following connection to the NSFnet. Column 2 accounts for time-varying factors within scientific disciplines common across all institution-pairs. Column 3 includes research inputs for both institutions i and j in the institution-pair d. Finally, Column 4 includes an institution-pair linear trend which accounts

---

[23] As previously discussed in Section 3.2, there can be no analysis on collaboration between broad scientific fields due to how the scientific field codes are assigned.

[24] Prior to the start of the NSFnet program, sixteen percent of all field-institution pairs had at least one collaboration.

[25] This strategy is similar to Agrawal and Goldfarb (2008) that studies how availability of BITNET impacted collaboration within electrical engineering.

for any changes in collaboration over time that may evolve linearly.

Results from estimating the change in the number of collaborations by QML Poisson are qualitatively similar. Shown in Table A.2, connection of institution-pairs to the national NSFnet did not significantly change total collaboration. At most, the estimates suggest that connections of institution-pairs to the national NSFnet led to small, imprecise declines in the likelihood of collaboration. However, it is difficult to argue the estimated effects are meaningful differences in the probability, or in the number of collaborations. Altogether, connection to the national NSFnet does not appear to have significantly altered collaborations between connected institutions.

5.2 Citation Pairs

Thus far, the results suggest that connection to the national NSFnet led to an increase in average citations per paper and that the documented change in citations is not due to changes in collaboration between connected institutions. Yet the source of the change in citations is unclear. If reductions in communication costs from the NSFnet created a "global village," then one would expect an increase in interdisciplinary citations. Alternatively, if reductions in communication costs allowed scientists within broad disciplines to further segregate within their scientific field and fostered scientific specialization, one would expect an increase in within-discipline citations.

To examine if the change in citations is driven by increased citations from within discipline, between disciplines, or both, I use the citation-pair data made available in the AC database. The data record the total citations made by one institution in some field to a another institution in a different (or the same) field.[26] I estimate the following equation to test if connection to the NSFnet increased inter-discipline citations, within-discipline citations, or both:

$$Citation_{dkt} = \omega NSFnetBoth_{dt} + \delta(NSFnetBoth_{dt} \times WithinField_k) + \eta_d + \theta_k + \varepsilon_{dkt} \quad (5)$$

where k denotes the scientific field pair of field n citing field m. The dependent variable,

---

[26] Using the earlier example, Stanford biology citing Harvard medicine in 1983.

$Citation_{dkt}$, is an indicator if there are *any* citations sent by institution i to institution j in institution (d) and field (k) pairs. Here, *NSFnetBoth* takes the value of one when both institutions in a pair are connected to the NSFnet in year t. *WithinField* takes the value of one when both institutions in the same pair belong to the same scientific discipline. Institution-pair fixed effects (η_d) account for any time-invariant factors that may influence citations sent between institution pairs. Discipline fixed effects (θ_k) account for time-invariant factors that may influence citations sent between pairs of broad scientific disciplines. Standard errors are clustered at the institution-pair level.

In Equation 5, ω represents the average change in the probability of citations between two universities across dissimilar disciplines following connection to the national NSFnet holding constant differences in institution-discipline pairs. The change in the probability of citations between two universities within scientific disciplines following connection to the national NSFnet is represented by δ. Then, the net effect of connection to the NSFnet between pairs of universities within the same scientific discipline is the linear combination of $\omega + \delta$.

Table 5 displays the result of estimating Equation 5 by a linear probability model. Column 1 of Table 5 presents the baseline estimates of the change in the probability of citations both between and within scientific disciplines. Column 2 accounts for time-varying factors that may influence citations between pairs of fields common across all institution-pairs. Column 3 includes research inputs for each institution in institution-pair d. Finally, Column 4 includes institution-pair specific linear trends. The estimates suggest that connection of institution-pairs to the NSFnet led to a three percent decline in the probability of interdisciplinary citations and an eight percent increase in the probability of within field citations, on average. Taken together, the estimated effect of national NSFnet connection between institution-pairs is robust across specifications.

One may be concerned, however, that some fields are closely connected and simply splitting fields by within field and between field may mask more detailed interactions between academic disciplines. To address this, I total the number of citations from one field to all fields prior to 1985 and create a ranked list of the most cited fields by an individual field.[27] Table A.4 provides an example of the pre-NSFnet citation ranking for Physics. Unsurpris-

---

[27] For a more detailed discussion of the construction of citation rankings, see Section A.2.3

ingly, the most cited field for Physics is Physics. Importantly, a similar pattern emerges for *all* disciplines: for each scientific field the most cited field is always itself. For Physics, the other four fields within the top five cited fields are Chemistry, Astronomy, Biology, and Medicine.

I estimate if the effect of the concurrent NSFnet connection between institutions differs by ranks of scientific disciplines. I modify Equation 5 to instead include an interaction between *NSFnetBoth* and indicators for field-pair rankings. Figure 13 presents the estimates and largely supports the findings presented in Table 5. For the two most cited fields, NSFnet connection increased the probability of citation with the most pronounced effect within field. The effect of the NSFnet on the probability of citation declines with distance from the most cited field to least cited field.

Additionally, connection to the NSFnet may alter citations between institutions along the intensive margin. Of the 11,204,680 institution-field-pair observations, 76 percent record no citation and of the non-zero citations, the median amount of citations is two.[28] Estimating the change in the total number of citations by QML Poisson estimation provides similar results. The estimates shown in Table A.3 suggest an, on average, eight percent decline in citations between disciplines and a seven percent increase in citations within disciplines. Using Poisson estimation, Figure A.1 echoes the same conclusion when estimating the changes in total number of citations between fields by rank.

All together, connection to the early internet seems to have increased within-discipline citations while reducing interdisciplinary citations. Taken with the negligible decline in the likelihood of collaboration, the national NSFnet appears to have fostered further specialization within scientific disciplines and reduced the likelihood of interdisciplinary citations between connected institutions. The results are consistent with theoretical predictions by Rosenblat and Mobius (2004): reductions in communication costs led researchers to further specialize within individual scientific fields while reducing interdisciplinary work.

---

[28] The average amount of pre-NSFnet citations made within-field and between-field are 1.19 and 0.1, respectively.

6 Conclusion

The NSFnet set the standard for large-scale computer networking services by allowing users to access e-mail, transfer files, and browse the World Wide Web on a reliable, high-speed network. The overall findings suggest that a decline in communication costs due to the national NSFnet bolstered scientific output academic institutions by increasing citations per paper roughly ten percent. Exploring potential mechanisms, the paper provides evidence that reductions in communication costs due to NSFnet connection was driven by a change in citation patterns between connected institutions. Scientific disciplines within institutions became further specialized as a result of national NSFnet connection. At the same time, I do not find evidence that the net effect from the NSFnet was driven by changes in collaborations between connected institutions.

This paper contributes to the literature by offering empirical evidence for theoretical predictions of communication costs in science. Reductions in communication costs allow individuals to search for, screen out, and curate information aligned with their preferences. Van Alstyne and Brynjolfsson (1996), Rosenblat and Mobius (2004), and Van Alstyne and Brynjolfsson (2005) hypothesize that reductions in communication costs may provide scientific researchers with better access to previously untapped knowledge and enable scholars to specialize further within their own discipline. The results of this paper provide evidence for these predictions by showing that connection to the national NSFnet led to increased citations between institutions, within scientific disciplines.

Furthermore, this paper also adds to the literature that studies early information and communication technologies in science by documenting how interactions between scientific disciplines changed following connection to an early version of the Internet. Previous work by Agrawal and Goldfarb (2008) and Ding et al. (2010) show that early computer network connectivity increased interactions between institutions within individual scientific disciplines. In this paper, I show that connection to the national NSFnet led to increased citations per paper and altered the composition of citations between institutions. These results expand the literature by documenting that internet connectivity altered interactions both within and between scientific disciplines.

Finally, the findings of this paper help to provide a better understanding of the effects of communication costs and ICTs on interdisciplinary research. Internet connectivity transformed science by reducing costs to access new ideas, connect with distant colleagues, and work contemporaneously on complex projects. Reductions in communications costs, combined with incentives faced by individual scholars, have made interdisciplinary research more challenging. Put differently, reductions in communication costs led researchers to forgo interdisciplinary research in favor of heightened specialization.

Innovation and new discoveries are often found through the combination of research, methods, and ideas across disciplines. As a result, policy makers, agencies that support fundamental research and education, and individual universities increasingly emphasize the need for collaboration across scientific fields. However as disciplines have further specialized, the incentive to engage in interdisciplinary research has diminished. In fact, engaging in interdisciplinary research is viewed by many scholars as an impediment toward funding, hiring, and promotion (National Academy of Sciences 2005).

To overcome these difficulties associated with connecting scientific fields, nearly all major research institutions have established offices to promote and fund interdisciplinary research through policies like interdisciplinary cluster hiring, grants, and a greater emphasis on interdisciplinary research for tenure considerations (Klein and Falk-Krzesinski 2017). Although policies to bolster research across disciplines are increasingly common, there is little causal evidence of the effectiveness of such policies on interdisciplinary research (Leahey and Barringer 2020). As ICTs become steadily more ubiquitous and disciplines further specialize, the effectiveness of policies aimed to support interdisciplinary research remains and important empirical question.

References

Adams, James D., Grant C. Black, J. Roger Clemmons, and Paula E. Stephan. 2005. “Scientific teams and institutional collaborations: Evidence from U.S. universities, 1981-1999”. *Research Policy* 34 (3): 259–285. ISSN: 00487333. doi:10.1016/j.respol.2005.01.014.

Adams, James D., and J. Roger Clemmons. 2008. “The NBER-Rensselaer Scientific Papers Database: Form, Nature, and Function”. Cambridge, MA. doi:10.3386/w14575.

— . 2011. “The role of search in university productivity: Inside, outside, and interdisciplinary dimensions”. *Industrial and Corporate Change* 20 (1): 215–251. ISSN: 09606491. doi:10.1093/icc/dtq071.

Agrawal, Ajay, and Avi Goldfarb. 2008. “Restructuring research: Communication costs and the democratization of university innovation”. *American Economic Review* 98 (4): 1578–1590. ISSN: 00028282. doi:10.1257/aer.98.4.1578.

Akerman, Anders, Ingvil Gaarder, and Magne Mogstad. 2015. “The Skill Complementarity of Broadband Internet”. *Quarterly Journal of Economics* 130 (4): 1781–1824. doi:10.1093/qje/qjv028.Advance.

Berkes, Enrico, and Peter Nencka. 2019. “Knowledge Access: The Effects of Carnegie Libraries on Innovation”.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. “How much should we trust differences-in-differences estimates?”, 119 (1): 249–275. ISSN: 00335533. doi:10.1162/003355304772839588.

Biasi, Barbara, and Petra Moser. 2020. “Effects of Copyrights on Science: Evidence from the WWII Book Republication Program”. *American Economic Journal: Microeconomics* Forthcoming.

Bloom, Nicholas, Luis Garicano, Raffaella Sadun, and John Van Reenen. 2014. “The Distinct Effects of Information Technology and Communication Technology on Firm Organization”. *Management Science* 60 (12): 2859–2885. ISSN: 0025-1909. doi:10.1287/mnsc.2014.2013.

Bloom, Nicholas, Raffaella Sadun, and John Van Reenen. 2012. “Americans Do IT better: US Multinationals and the Productivity Miracle”. *American Economic Review* 102 (1): 167–201. ISSN: 00028282. doi:10.1257/aer.102.1.167. arXiv: 00368075.

Chetty, Raj, Adam Looney, and Kory Kroft. 2009. “Salience and taxation: Theory and evidence”. *American Economic Review* 99 (4): 1145–1177. ISSN: 00028282. doi:10.1257/aer.99.4.1145.

Claffy, Kimberly C., George C. Polyzos, and Hans-Werner Braun. 1993. “Traffic Characteristics of the T1 NSFNET Backbone”. *IEEE*: 885–892.

Ding, Waverly W., Sharon G. Levin, Paula E. Stephan, and Anne E. Winkler. 2010. “The Impact of Information Technology on Academic Scientists’ Productivity and Collaboration Patterns”. *Management Science* 56 (9): 1439–1461. ISSN: 0025-1909. doi:10.1287/mnsc.1100.1195.

Forman, Chris, and Nicolas Van Zeebroeck. 2012. “From Wires to Partners: How the Internet Has Fostered R&D Collaborations within Firms”. *Management Science* 58 (8): 1549–1568. ISSN: 17437563. doi:10.1111/ajco.12822.

Frazer, Karen D.; 1995. “NSFNET : A Partnership for High-Speed Networking Final Report”. *Merit Network* 6 (4): 44.

Furman, Jeffrey L, Markus Nagler, and Martin Watzinger. 2018. *Disclosure and Subsequent Innovation: Evidence from the Patent Depository Library Program*, NBER Working Paper Series 24660. NBER. doi:10.3386/w24660.

Goldfarb, Avi, and Catherine Tucker. 2019. “Digital Economics”. *Journal of Economic Literature* 57 (1): 3–43.

Gross, Daniel P. 2019. *The Consequences of Invention Secrecy: Evidence From The USPTO Patent Secrecy Program in World War II*, NBER Working Paper Series 25545. NBER.

Gurbaxani, Vijay. 1990. “Diffusion in computing networks: the case of BITNET”. *Communications of the ACM* 33 (12): 65–75.

Hafner, Katie, and Matthew Lyon. 1998. *Where Wizards Stay Up Late: The Origins of the Internet*. Simon & Schuster.

Hjort, Jonas, and Jonas Poulsen. 2019. "The arrival of fast internet and employment in Africa". *American Economic Review* 109 (3): 1032–1079. ISSN: 0002-8282. doi:10.3386/w23582.

Horton, John J., and Richard J. Zeckhauser. 2016. *Owning, Using, and Renting: Some Simple Economics of the "Sharing Economy"*, Working Paper Series 22029. NBER.

Iaria, Alessandro, Carlo Schwarz, and Fabian Waldinger. 2018. "Frontier knowledge and scientific production: Evidence from the collapse of international science". *Quarterly Journal of Economics* 133 (2): 927–991. ISSN: 15314650. doi:10.1093/qje/qjx046.

Jones, Benjamin F. 2009. "The Burden of Knowledge and the "Death of the Renaissance Man": Is Innovation Getting Harder?" *Review of Economic Studies* 76:283–317.

Jones, Benjamin F., Stefan Wuchty, and Brian Uzzi. 2008. "Multi-University Research Teams: Shifting Impact, Geography, and Stratification in Science". *Science* 322:1259–1262. ISSN: 0164-0259. doi:10.1177/107554709401500403.

Klein, Julie Thompson, and Holly J. Falk-Krzesinski. 2017. "Interdisciplinary and collaborative work: Framing promotion and tenure practices and policies". *Research Policy* 46 (6): 1055–1061. ISSN: 0048-7333. doi:https://doi.org/10.1016/j.respol.2017.03.001.

Kuhn, Peter, and Hani Mansour. 2014. "Is Internet Job Search Still Ineffective?" *Economic Journal* 124 (581): 1213–1233. ISSN: 14680297. doi:10.1111/ecoj.12119.

LaQuey, Tracy L. 1990. *The User's Directory of Computer Networks.* 247–380. Digital Press. ISBN: 1-55558-047-5.

Leahey, Erin, and Sondra N. Barringer. 2020. "Universities' commitment to interdisciplinary research: To what end?" *Research Policy* 49 (2): 103910. ISSN: 0048-7333. doi:https://doi.org/10.1016/j.respol.2019.103910.

Mokyr, Joel. 2002. *The gifts of Athena: Historical origins of the knowledge economy.* Princeton University Press.

— . 2005. "The Intellectual Origins of Modern Economic Growth". *The Journal of Economic History* 65 (2): 285–351.

Murray, Fiona, and Scott Stern. 2007. "Do formal intellectual property rights hinder the free flow of scientific knowledge?. An empirical test of the anti-commons hypothesis". *Journal of Economic Behavior and Organization* 63 (4): 648–687. ISSN: 01672681. doi:10.1016/j.jebo.2006.05.017.

National Academy of Sciences. 2005. *Facilitating Interdisciplinary Research*. Washington, DC: The National Academies Press. ISBN: 978-0-309-09435-1. doi:10.17226/11153.

Rosenblat, Tanya S., and Markus M. Mobius. 2004. "Getting closer or drifting apart?" *Quarterly Journal of Economics* 119 (3): 971–1009. ISSN: 00335533. doi:10.1162/0033553041502199.

Schmidheiny, Kurt, and Sebastian Siegloch. 2019. "On Event Study Designs and Distributed-Lag Models : Equivalence , Generalization and Practical Implications".

Singh, Jasjit, and Lee Fleming. 2010. "Lone inventors as sources of breakthroughs: Myth or reality?" *Management Science* 56 (1): 41–56. ISSN: 00251909. doi:10.1287/mnsc.1090.1072.

Teodoridis, Florenta. 2018. "Understanding team knowledge production: The interrelated roles of technology and expertise". *Management Science* 64 (8). ISSN: 15265501. doi:10.1287/mnsc.2017.2789.

Van Alstyne, Marshall, and Erik Brynjolfsson. 1996. "Could the Internet Balkanize Science ?" *Science* 274 (5292): 1479–1480.

— . 2005. "Global Village or Cyber-Balkans? Modeling and Measuring the Integration of Electronic Communities". *Management Science* 51 (6): 851–868. ISSN: 0025-1909. doi:10.1287/mnsc.1050.0363.

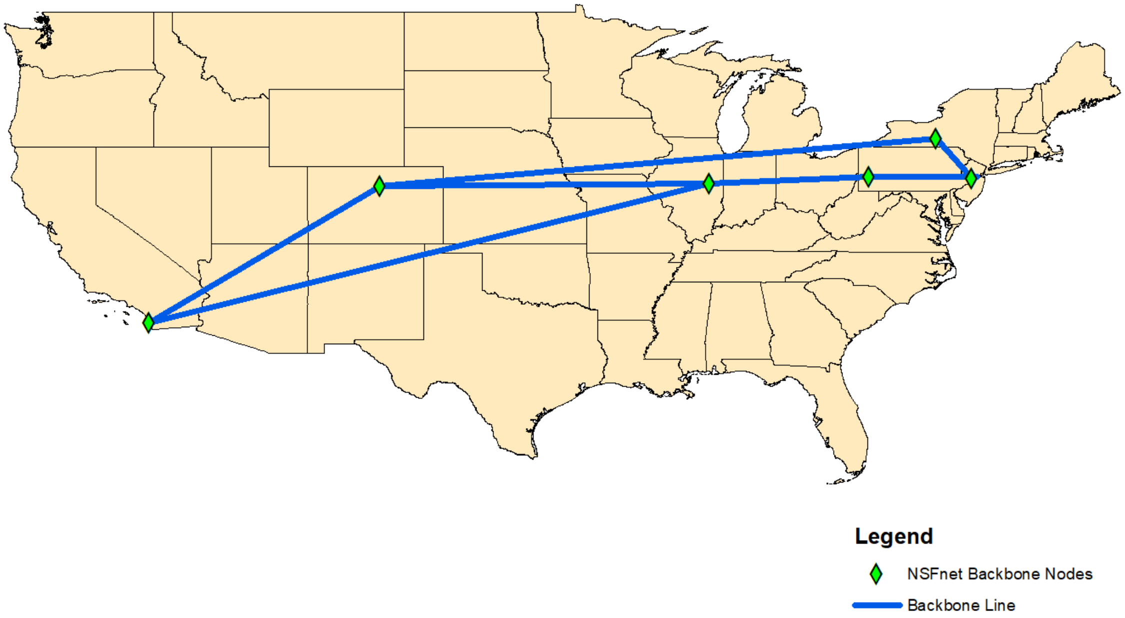
Weitzman, Martin L. 1998. "Recombinant growth". *The Quarterly Journal of Economics* 113 (2): 331–360.

Wuchty, Stefan, Benjamin F Jones, and Brian Uzzi. 2007. "The increasing dominance of teams in production of knowledge". *Science* 316 (5827): 1036–1039.

Figure 1: The NSF Supercomputers

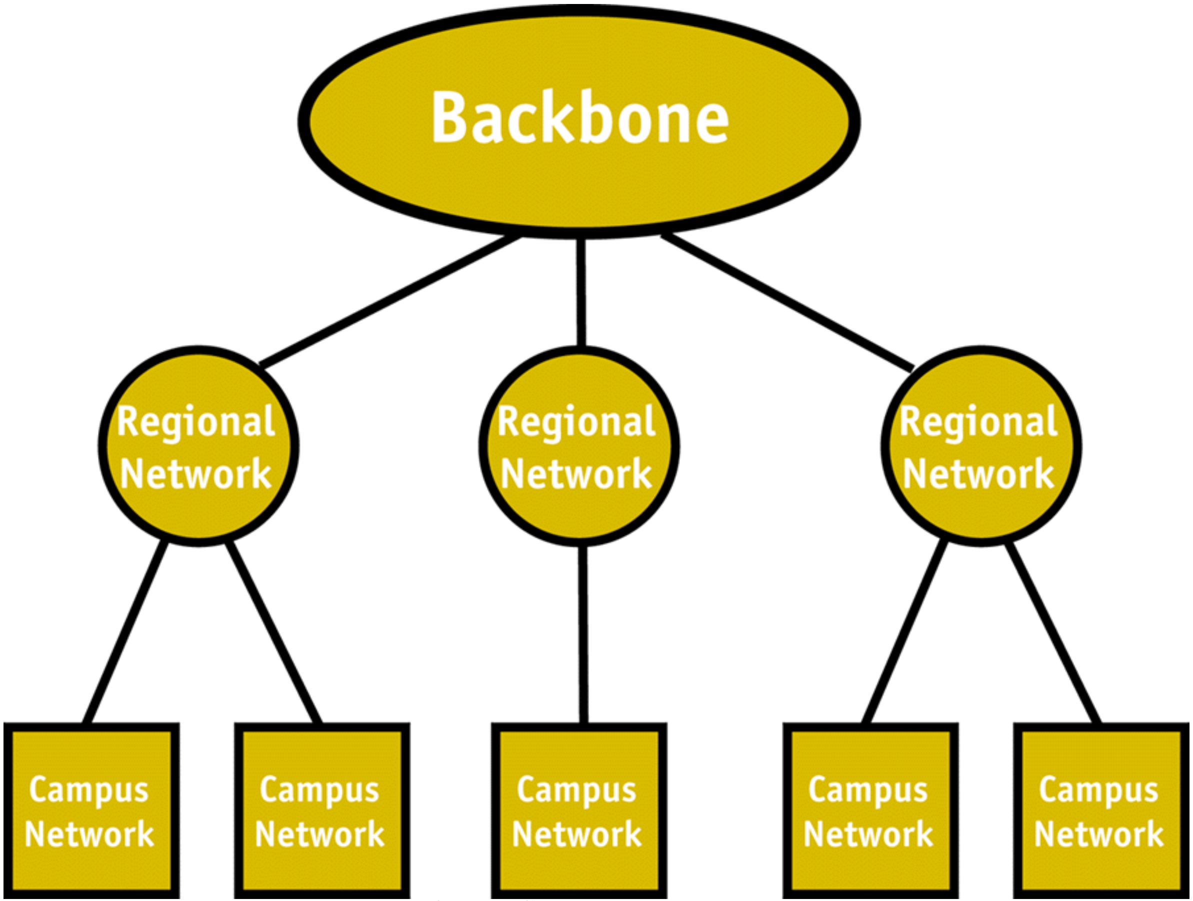
Note: Figure shows the location of the five NSF sponsored super-computing centers plus NCAR. The NSFnet began as a resource-sharing network to provide distant institutions access to supercomputing centers. Each super-computing center formed the original NSFnet national internet backbone. *Source: The User's Directory to Computer Networks*

Figure 2: The NSFnet, 1986



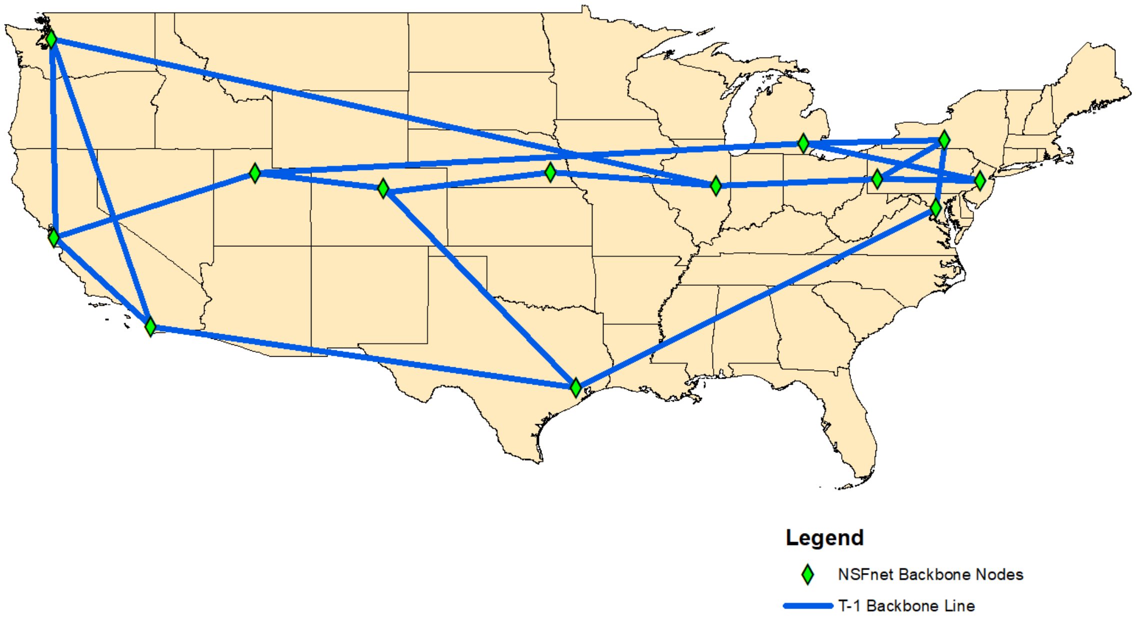
Note: Figure shows the initial NSFnet backbone network operational in July of 1986. The original network provided access to federally funded super-computing centers to campuses nearby. Shortly after its launch, MERIT took over management of the NSFnet and quickly expanded the network. *Source: The User's Directory to Computer Networks*

Figure 3: The NSFnet Three-Tiered Architecture



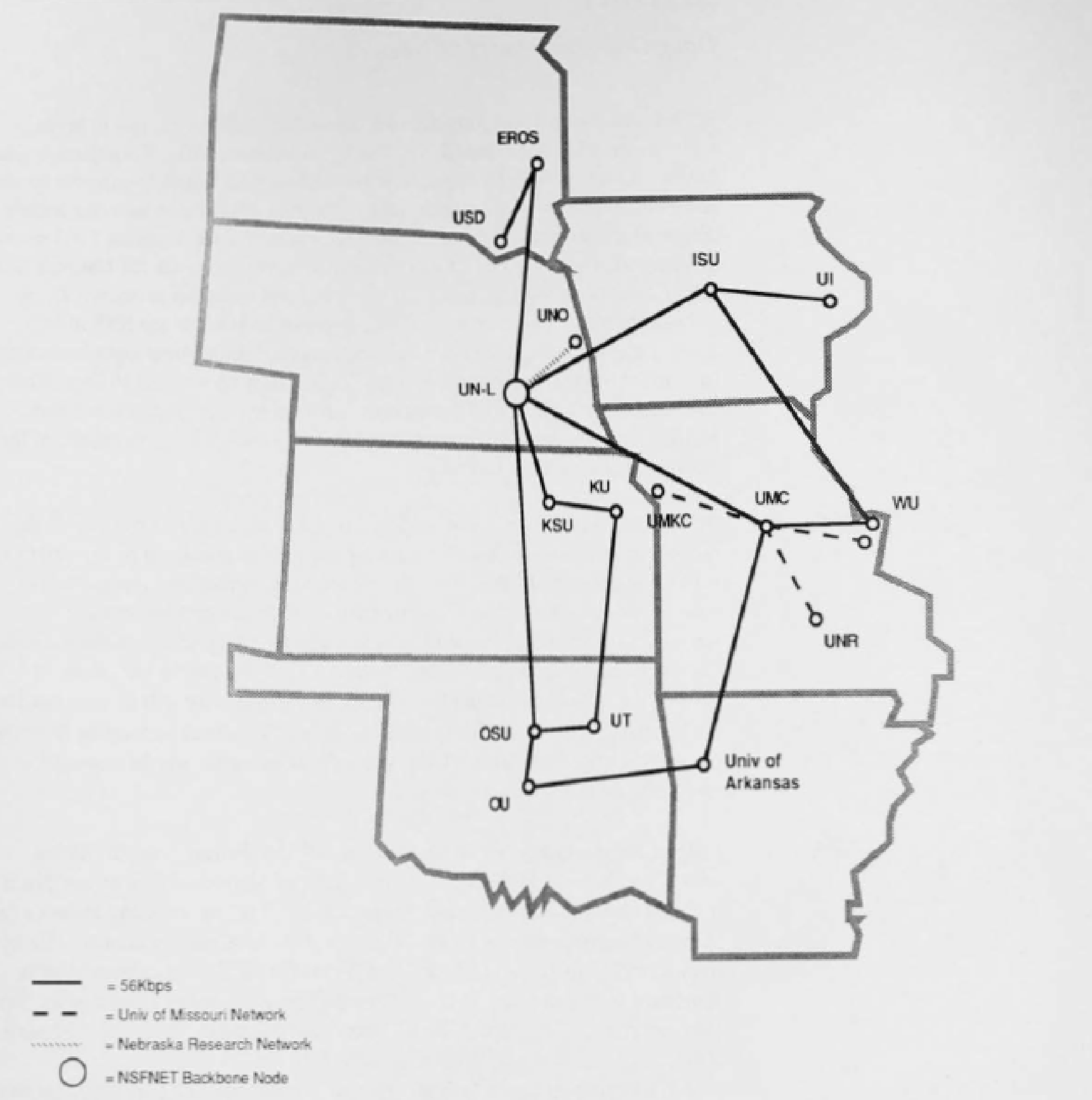
Note: Figure shows the architecture of the NSFnet. Rather than linking each individual participant to the network, the NSFnet connected regional computer networks to the national internet backbone. The NSFnet program managed the connection of regional networks to the backbone, while connecting individual universities was left up to the managers of regional networks. *Source*: `https://www.livinginternet.com/doc/merit.edu/government.html`

Figure 4: The NSFnet, 1990



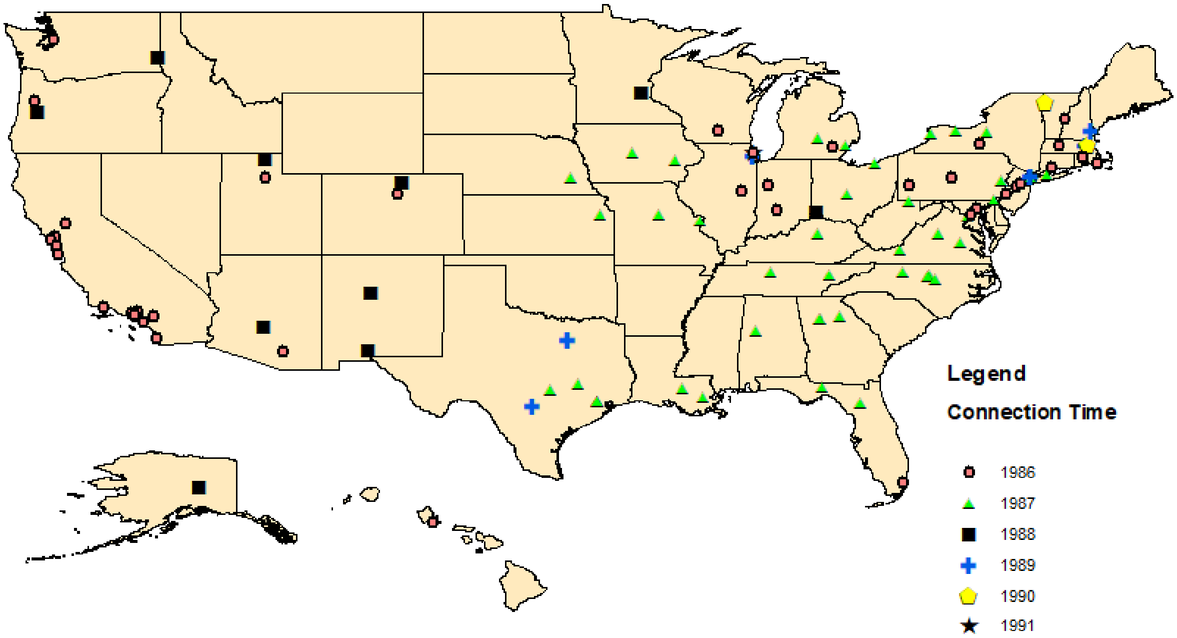
Note: Figure shows the NSFnet at the time of *The User's Guide to Computer Networks*. The thirteen primary backbone nodes are represented in green (diamond). Each backbone node typically acted as a host to at least one regional network comprised of several institutions.

Figure 5: Example of *The User's Directory of Computer Networks*



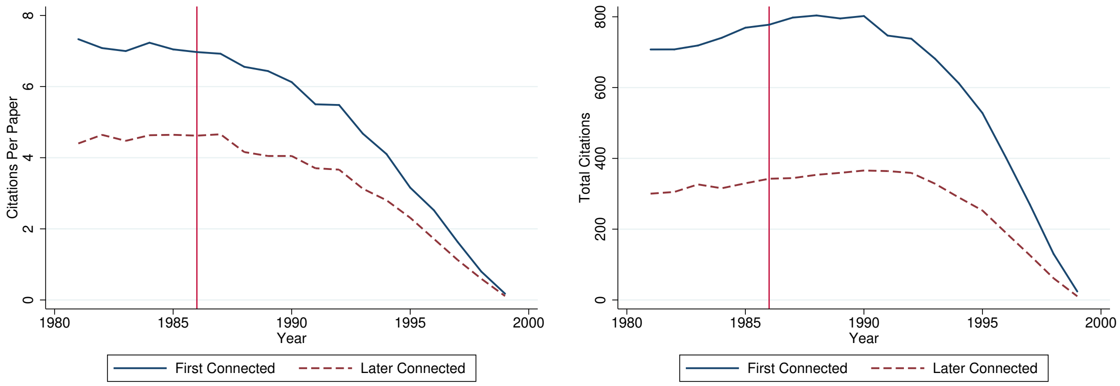
The figure displays the map of MIDnet, a NSFnet regional network founded in spring of 1985 in order to provide researchers at mid-western institutions access to super-computing resources. *Source: The User's Directory to Computer Networks*

Figure 6: Connection Times to the NSFnet

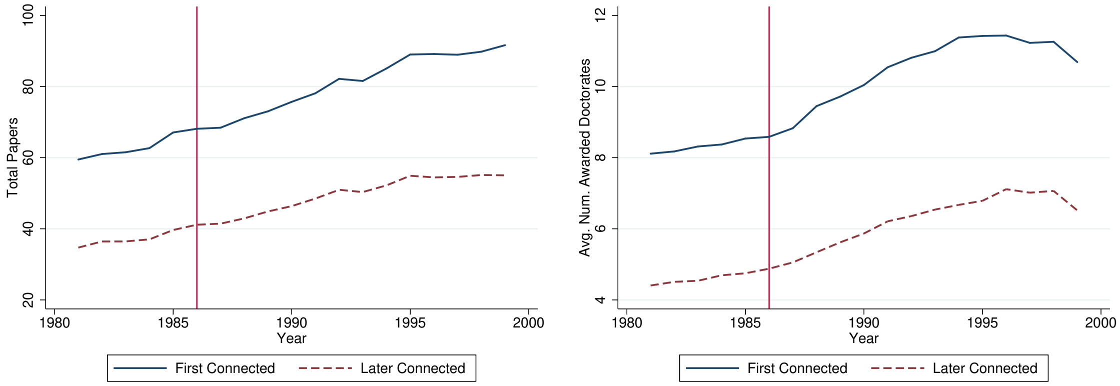


Note: Figure shows the connection years of individual educational institutions to the NSFnet. Each institution was part of one or several regional computer-networks connected to the NSFnet. An individual institution is assigned the earliest year of regional NSFnet connection. *Source: The User's Directory to Computer Networks*

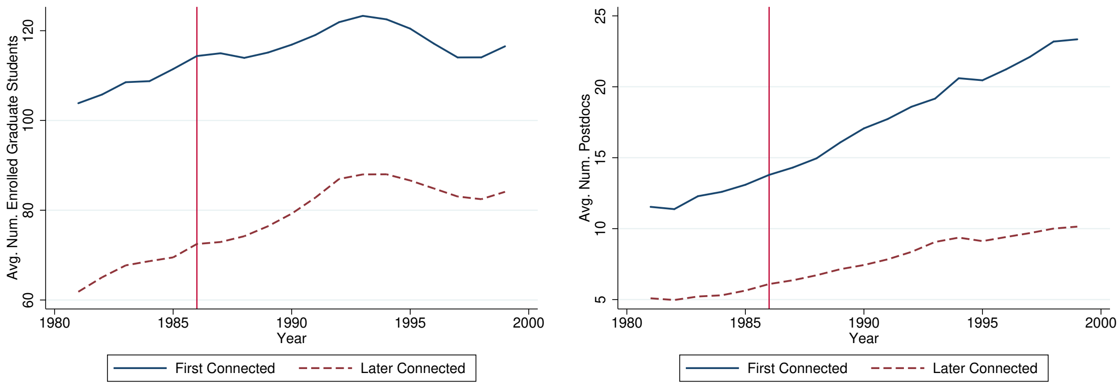
Figure 7: Trends in Observables by Connection Group – CASPAR20 Variables


(a) Citations Per Paper

(b) Total Citations


(c) Total Papers

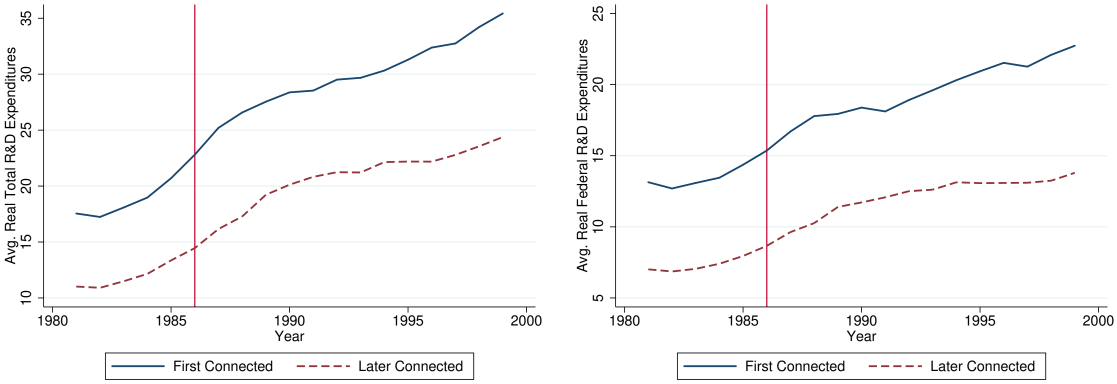
(d) Awarded Doctorates


(e) Enrolled Graduate Students

(f) Number of Postdocs

Note: Figures show trends in observable variables collected at the CASPAR20-level. Annual means are computed by taking the annual average of each variable by connection group. *Source: author's calculations.*

Figure 8: Trends in Observables by Connection Group – CASPAR12 Variables

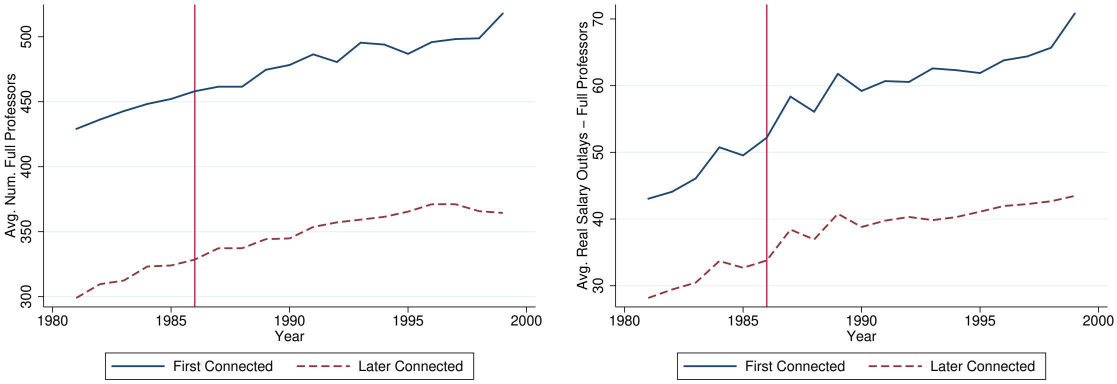


(a) Total Real R&D Expenditures

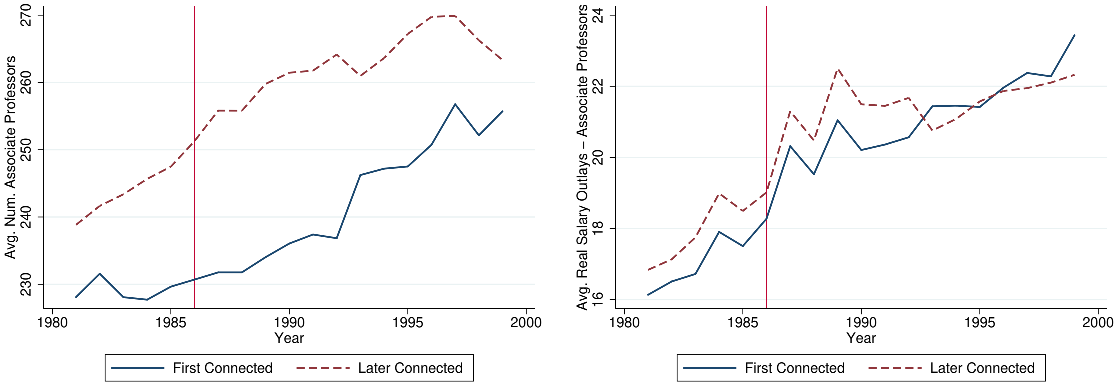
(b) Federal Real R&D Expenditures

Note: Figures show trends in observable variables collected at the CASPAR12-level. Annual means are computed by taking the annual average of each variable by connection group. *Source: author's calculations.*

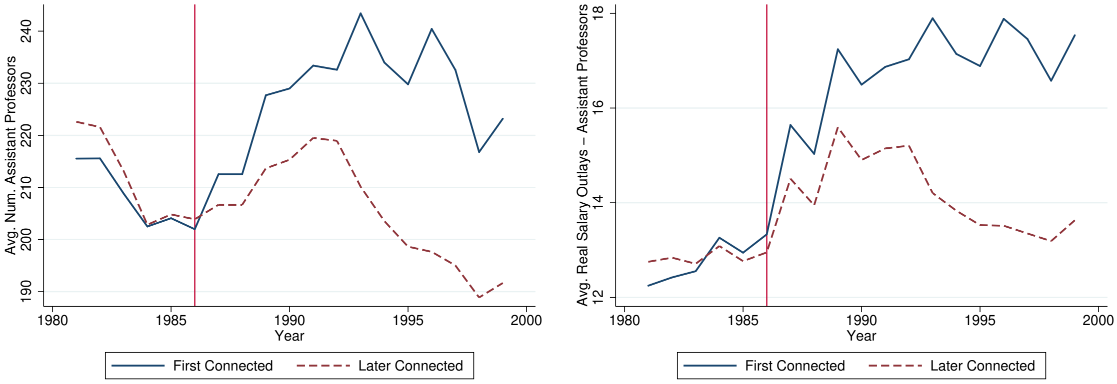
Figure 9: Trends in Observables by Connection Group – Institution-Level Variables


(a) Number of Faculty - Professors

(b) Total Salary Outlays - Professors


(c) Number of Faculty - Associate Profs.

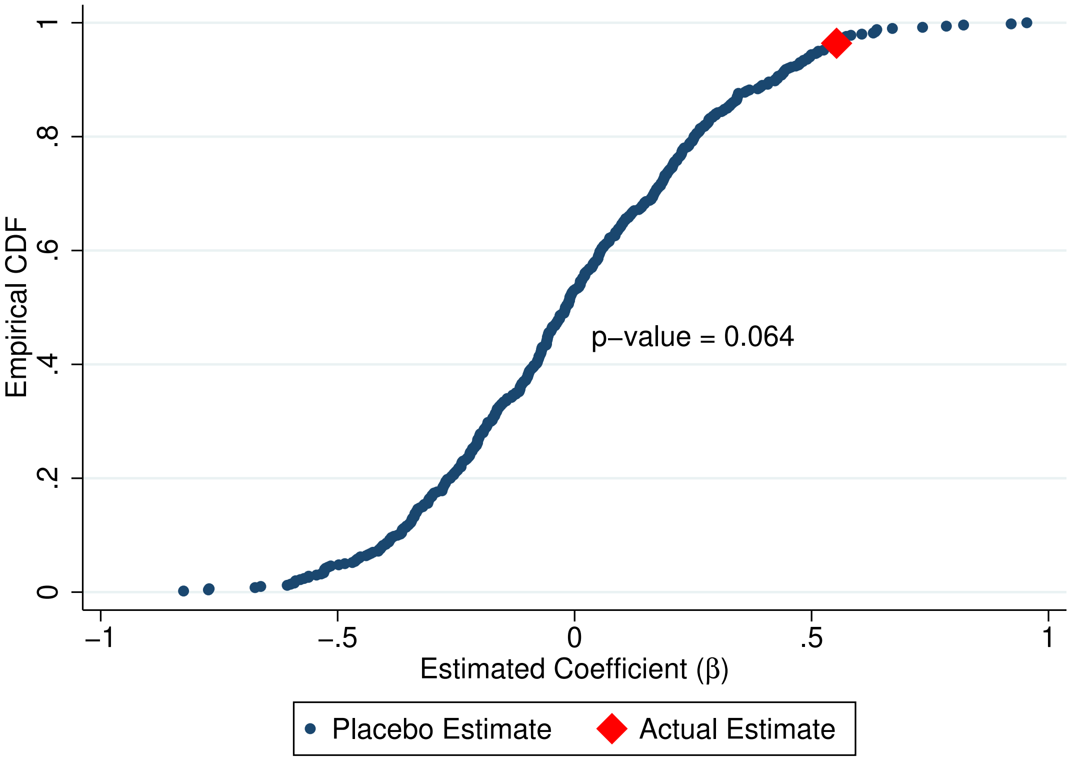
(d) Total Salary Outlays - Associate Profs.


(e) Number of Faculty - Assistant Profs.

(f) Total Salary Outlays - Assistant Profs.

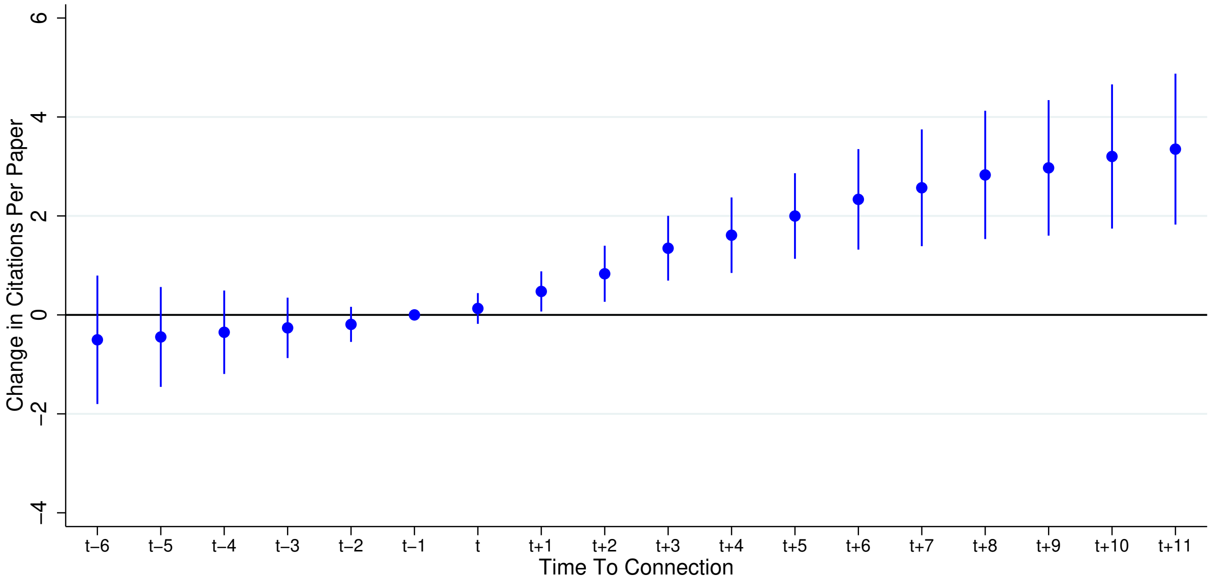
Note: Figures show trends in observable variables collected at the individual institution-level. Annual means are computed by taking the annual average of each variable by connection group. *Source: author's calculations.*

Figure 10: Distribution of Placebo Estimates



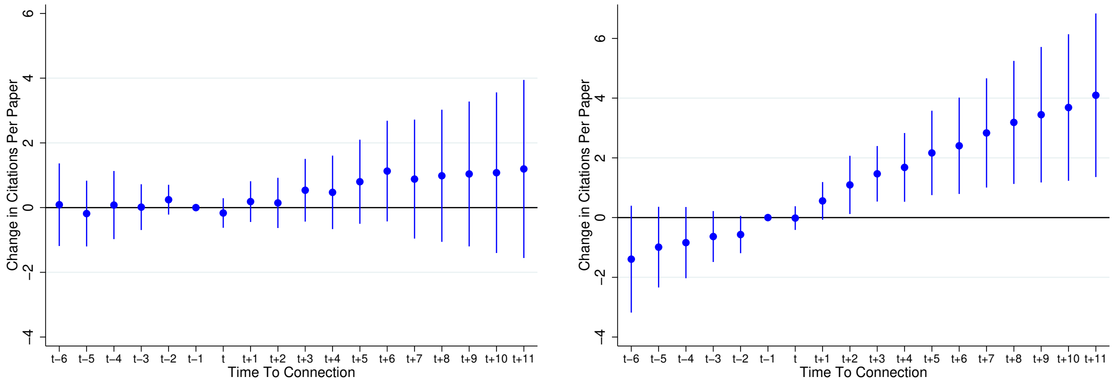
Note: Figure plots the empirical distribution of placebo estimates for citations per paper. The empirical distribution is constructed from 500 placebo estimates of β that employ the specification in Column 1 of Table 2. The red diamond shows the actual estimated effect of national NSFnet connection on citations per paper reported in column 1 of Table 2.

Figure 11: Event Study Estimates

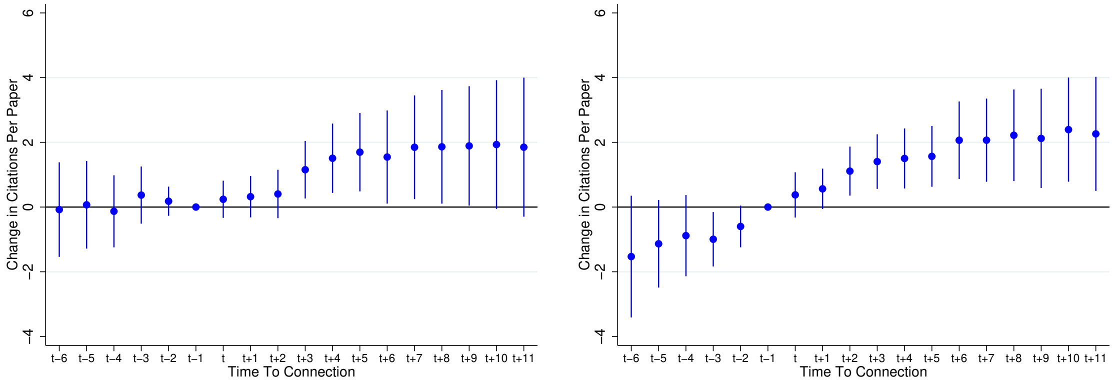


Note: Figure plots the estimated β_ts from Eq. 3. The omitted year is the year prior to national NSFnet connection (i.e. $t-1$). The 95% confidence intervals are constructed using institution-clustered standard errors.

Figure 12: Heterogeneity by Pre-NSFnet Average Salary


(a) Lowest Quartile

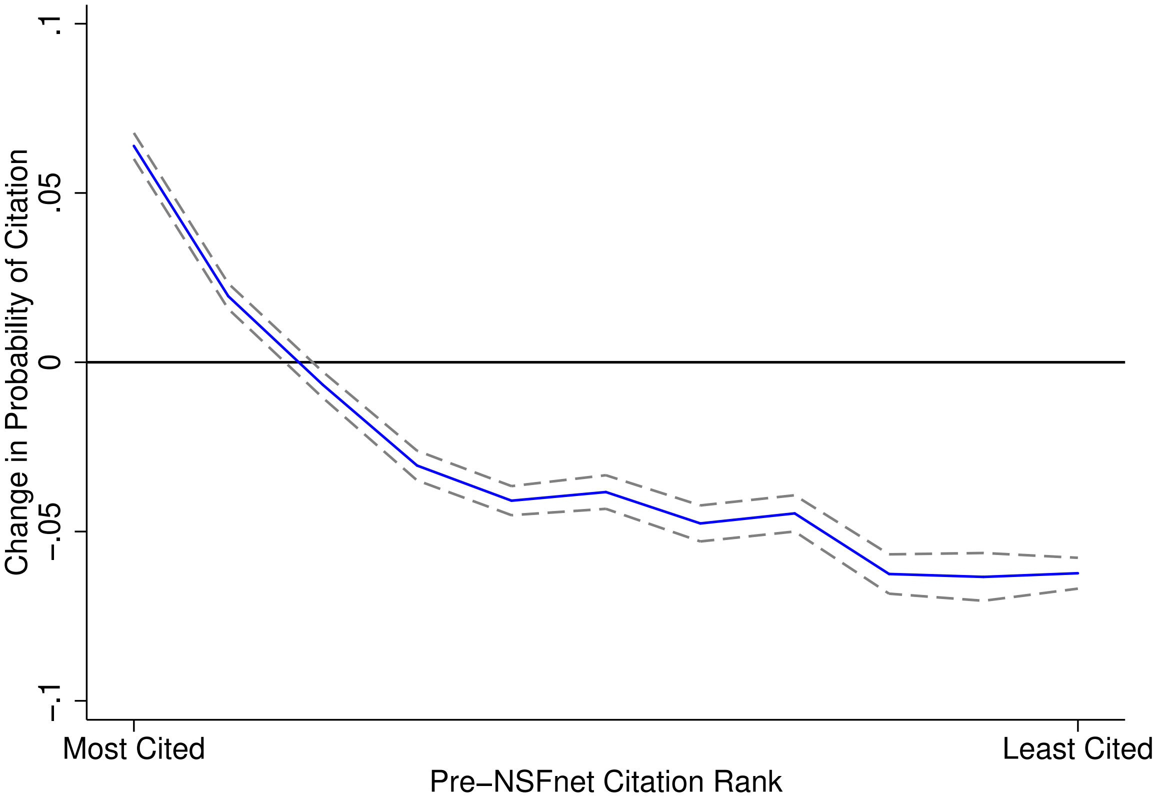
(b) 2nd Quartile


(c) 3rd Quartile

(d) Highest Quartile

Note: Figure plots the estimated β_t's by quartile of pre-NSFnet average faculty salary. The omitted year is the year prior to national NSFnet connection (i.e. $t-1$). The 95% confidence intervals are constructed using institution-clustered standard errors.

Figure 13: Change in Probability of Citation by Field Citation Rank



Note: Figure shows the estimated change in the probability of citing another institution across differing scientific fields following institution pair's connection to the NSFnet. It presents estimates and 95% confidence intervals in blue and gray lines, respectively. The pre-NSFnet citation rank orders the amount of citations made by an individual scientific discipline prior to the NSFnet program. The highest rank, or most cited field, is always within discipline (e.g. Physics citing Physics). The figure indicates that, following NSfnet connection, the likelihood of citing the two most cited fields increases while the likelihood of citing all other fields declines. Standard errors are clustered at the institution-pair level.

Table 1: Pre-NSFnet Summary Statistics

Variable	(1) All Institutions	(2) First Connected	(3) Later Connected	(4) Difference in Means (*t-2*)	(5) Difference in Trends (*t-2*)-(*t-3*)
Panel (A) CASPAR20 Variables					
Citations Per Paper	5.65	7.14	4.56	4.96***	0.78
	[6.72]	[7.42]	[5.93]	(0.77)	(0.97)
Total Citations	489.70	728.82	315.42	647.02***	17.36
	[1,618.31]	[2,168.46]	[1,014.33]	(92.31)	(21.07)
Total Papers	47.61	62.36	36.86	42.21***	-0.86
	[109.00]	[134.17]	[84.54]	(10.75)	(1.74)
Doctorates Awarded	6.13	8.30	4.58	6.67***	-0.34
	[10.92]	[12.80]	[9.03]	(1.17)	(0.34)
Enrolled Graduate Students	83.66	107.67	66.55	83.36***	-0.67
	[131.21]	[161.02]	[101.53]	(13.24)	(1.65)
Number of Postdocs	8.13	12.18	5.24	11.19***	0.10
	[29.35]	[40.25]	[17.36]	(1.89)	(0.29)
N	10,650	4,490	6,160	2,110	2,110
Panel (B) CASPAR12 Variables					
Federal Real R&D Expenditures	7.48	10.57	5.26	7.37***	0.15
	[17.34]	[21.00]	[13.71]	(1.51)	(0.21)
Total Real R&D Expenditures	11.20	14.82	8.59	10.92***	0.27
	[23.46]	[27.82]	[19.34]	(2.21)	(0.34)
N	6,460	2,700	3,760	1,292	1,292
Panel (C) Institution-Level Variables					
Number of Full Profs.	363.45	441.02	309.27	296.64***	5.61
	[223.53]	[224.96]	[206.20]	(34.51)	(5.40)
Number of Associate Profs.	235.47	228.67	240.22	0.34	-6.40**
	[134.60]	[131.23]	[136.90]	(20.67)	(3.19)
Number of Assistant Profs.	209.79	208.96	210.37	33.16**	-4.31
	[116.84]	[103.41]	[125.53]	(15.49)	(3.18)
Real Total Salaries Full Profs.	37.11	46.63	30.47	33.02***	3.36***
	[23.60]	[23.93]	[20.97]	(3.89)	(0.84)
Real Total Salaries Associate Profs.	17.33	16.93	17.62	-2.31	-0.57*
	[10.16]	[9.89]	[10.35]	(1.60)	(0.32)
Real Total Salaries Assistant Profs.	12.67	12.67	12.68	1.20	-0.05
	[7.40]	[6.31]	[8.09]	(1.01)	(0.24)
N	535	220	315	107	107

* $p < .10$, ** $p < .05$, *** $p < .01$

Notes: Standard Deviations reported in brackets. Cluster-robust standard errors at the institution-level are reported in parentheses. Column 1 presents the means across institutions prior to 1986, the beginning of the NSFnet program. Columns 2 and 3 present pre-NSFnet program means for first connected and later connected institutions, respectively. Column 4 presents a difference in means between first and later connected institutions two years prior to NSFnet connection, controlling for year effects. Column 5 presents a difference in trends between first and later connected institutions two years prior to NSFnet connection, controlling for year effects. Panels A, B, and C, contain variables reported at the institution-year-CASPAR20, institution-year-CASPAR12, and institution-year levels respectively. For a detailed breakdown of the NSF-CASPAR field codes see Section A.1.

Table 2: Change in Academic Output Following National NSFnet Connection

Dep. Variable	(1)	(2)	(3)	(4)	Pre-NSFnet $\bar{Y}$
Citations Per Paper	0.553***	0.565***	0.680***	0.217*	5.65
	(0.142)	(0.142)	(0.149)	(0.124)	[6.72]
Total Citations	110.010***	112.420***	166.986***	56.944***	489.70
	(21.140)	(21.540)	(36.074)	(13.472)	[1,618.31]
Total Papers	-2.033***	-2.113***	-0.353	0.024	47.61
	(0.741)	(0.750)	(0.875)	(0.672)	[109.00]
N	40,470	40,470	40,090	40,090	
Institution and Year FE	Y	Y	Y	Y	
CASPAR20 × Year FE	N	Y	Y	Y	
Time-Varying Controls	N	N	Y	Y	
Institution Linear Trend	N	N	N	Y	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: Table presents the estimated effect of national NSFnet connection on three measures of academic output. Column 1 presents a baseline estimate of Eq. 2 that includes institution and year fixed effects. Column 2 flexibly controls for field-specific effects that are constant across institutions. Column 3 includes time-varying controls to proxy for research inputs that may be correlated with both academic output and NSFnet connection. Column 4 includes an institution-specific linear time trend. The final column presents the pre-NSFnet connection mean of each dependant variable. Standard errors clustered at the institution level presented in parentheses and standard deviations are in brackets.

Table 3: Robustness of National NSFnet Connection

Dep. Variable	(1)	(2)	(3)
Citations Per Paper	0.564***	0.731***	0.688***
	(0.140)	(0.150)	(0.161)
Total Citations	141.492***	186.701***	167.896***
	(31.448)	(42.456)	(37.792)
Total Papers	0.362	-1.771*	-0.329
	(0.851)	(0.979)	(0.963)
N	40,090	40,090	36,993
Two-Year Lags of Controls	Y	N	N
Whole Counts	N	Y	N
BITNET Institutions	N	N	Y

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: Table presents the estimated effect of NSFnet connection. Each specification includes time-varying controls to proxy for research inputs as well as institution, year, and field-by-year fixed effects. Column 1 includes two-year lags of controls. Column 2 estimates Equation 2 with whole counts of citations per paper, total citations, and total papers as the dependant variables. Column 3 includes variation of BITNET connection timing from Agrawal and Goldfarb (2008). Standard errors clustered at the institution level are presented in parentheses.

Table 4: Change in Probability of Collaboration

	(1)	(2)	(3)	(4)
NSFnet Both	-0.003	-0.005**	-0.006**	0.000
	(0.003)	(0.003)	(0.003)	(0.003)
N	836,494	836,494	829,331	829,331
Institution Collaboration Pair and Year FE	Y	Y	Y	Y
CASPAR20-Pair × Year FE	N	Y	Y	Y
Time-Varying Controls	N	N	Y	Y
Institution Collaboration Pair Linear Trend	N	N	N	Y

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: Table presents estimates from a linear probability model that estimates Eq. 4. The dependent variable is equal to one if there is any collaboration between institutions i and j in field f during year t. *NSFnet Both* is the independent variable of interest that is equal to one when both institutions in pair d are connected to the national NSFnet. Column 1 is the baseline specification that includes institution-field-pair and year fixed effects. Column 2 captures the time-varying factors in collaboration within an individual scientific discipline that are common across all institution-pairs. Time-varying controls for each institution i and j in pair d are included in Column 3. Finally, Column 4 includes a linear time trend in institution-pair d. Institution-pair clustered standard errors are presented in parentheses.

Table 5: Change in Probability of Citation

	(1)	(2)	(3)	(4)
NSFnet Both (*Between* Fields)	-0.030***	-0.021***	-0.025***	-0.017***
	(0.001)	(0.001)	(0.001)	(0.001)
NSFnet Both (*Within* Fields)	0.080***	0.064***	0.066***	0.075***
	(0.001)	(0.002)	(0.002)	(0.002)
N	11,204,680	11,204,642	11,100,541	11,100,541
Institution-Citation-Pair and Year FE	Y	Y	Y	Y
CASPAR20-Pair × Year FE	N	Y	Y	Y
Time-Varying Controls	N	N	Y	Y
Institution-Citation-Pair Linear Trend	N	N	N	Y

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: Table presents estimates from a linear probability model that estimates Eq. 5. The dependent variable is equal to one if there is any citation between institutions i and j of institution-pair d within field-pair k during year t. *NSFnetBoth (Between Fields)* presents the estimates of ω, where the independent variable of interest is equal to one when both institutions in pair d are connected to the national NSFnet. *NSFnetBoth (Within Fields)* presents the estimates of the linear combination $\omega + \delta$, where the independent variable of interest is equal to one when both institutions in pair d are connected to the national NSFnet and are within the same field-pair k. Column 1 is the baseline specification that includes institution-field-pair and year fixed effects. Column 2 captures the time-varying factors in collaboration within an individual scientific discipline that are common across all institution-pairs. Time-varying controls for each institution i and j in pair d are included in Column 3. Finally, Column 4 includes a linear time trend in institution-pair d. Institution-pair clustered standard errors are presented in parentheses.

A Appendix

A.1 NSF CASPAR Code Hierarchy

The data assembled by Adams and Clemmons (2008) contain counts of an individual institution's publications and citations per year for 88 detailed Institute for Scientific Information (ISI) codes. Importantly, Adams and Clemmons (2008) map each detailed ISI code to a corresponding NSF-CASPAR field code.

The NSF-CASPAR field codes are used by the NSF to harmonize survey data among universities and the sciences. The NSF employs two primary definitions of the CASPAR field codes, one detailed and one general. The NSF-CASPAR12 field codes are a set of twelve general NSF-CASPAR field codes, while the NSF-CASPAR20 are a set of twenty detailed NSF-CASPAR field codes.

Table A.1: CASPAR Code Hierarchy

CASPAR12	CASPAR20	
Agriculture	Agriculture	
Astronomy	Astronomy	
Biology	Biology	
Chemistry	Chemistry	
Computer Sci.	Computer Sci.	
Economics	Economics	
Math and Stat.	Math and Stat.	
Medicine	Medicine	
Physics	Physics	
Psychology	Psychology	
Total Engineering	Aerospace Eng.	Chemical Eng.
	Civil Eng.	Electrical Eng.
	Industrial Eng.	Material Sci.
	Mechanical Eng.	Other Eng.
Total Earth Science	Earth Science	Oceanography

Table A.1 details the hierarchy of the CASPAR codes. The CASPAR12 codes nearly map the CASPAR20 codes one-to-one. The ultimate difference between hierarchies arises due to more detailed sub-fields of engineering and earth science captured by the CASPAR20 codes.

A.2 Alternate Estimation of Collaboration and Citations

A.2.1 Collaborations

Table A.2: Estimated Change in Collaboration

	(1)	(2)	(3)
NSFnet Both	-0.029	-0.021	-0.013
	(0.030)	(0.030)	(0.025)
N	836,494	836,494	829,331
Institution Collaboration Pair and Year FE	Y	Y	Y
CASPAR20 × Year FE	N	Y	Y
Time-Varying Controls	N	N	Y

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Table shows estimates from a QML Poisson model of the change in the number of collaborations between pairs of institutions following concurrent connection to the national NSFnet. The dependent variable is the number of collaborations between institutions i and j within field f during year t. *NSFnetBoth* is the independent variable of interest that is equal to one when both institutions of the pair d are connected to the national NSFnet. Institution-Pair clustered standard errors are presented in parentheses.

A.2.2 Citations

Table A.3: Estimated Change in Citations

	(1)	(2)	(3)
NSFnet Both (*Between* Fields)	0.105***	-0.078***	-0.066***
	(0.008)	(0.011)	(0.012)
NSFnet Both (*Within* Fields)	0.005	0.074***	0.095***
	(0.006)	(0.006)	(0.007)
N	11,204,680	10,999,151	10,897,043
Institution Collaboration Pair and Year FE	Y	Y	Y
CASPAR20 × Year FE	N	Y	Y
Time-Varying Controls	N	N	Y

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Table shows estimates from a QML Poisson model of the change in the number of citations between pairs of institutions following concurrent connection to the national NSFnet. The dependent variable is the number of citations between institutions i and j within field f during year t. *NSFnetBoth* is the independent variable of interest that is equal to one when both institutions of the pair d are connected to the national NSFnet. Institution-Pair clustered standard errors are presented in parentheses

A.2.3 Citation Ranking

This section details the construction of the citation rankings used to construct Figure 13. For each CASPAR20 scientific field, I total the amount of citations to all fields prior to the NSFnet program. Next, I sort each cited field by the total amount of citations made by the individual citing field. In the event of a tie (two or more fields were cited the same amount prior to the NSFnet), I assign the higher rank to each field.[29]

Table A.4: Physics to Other Fields Citation Ranking

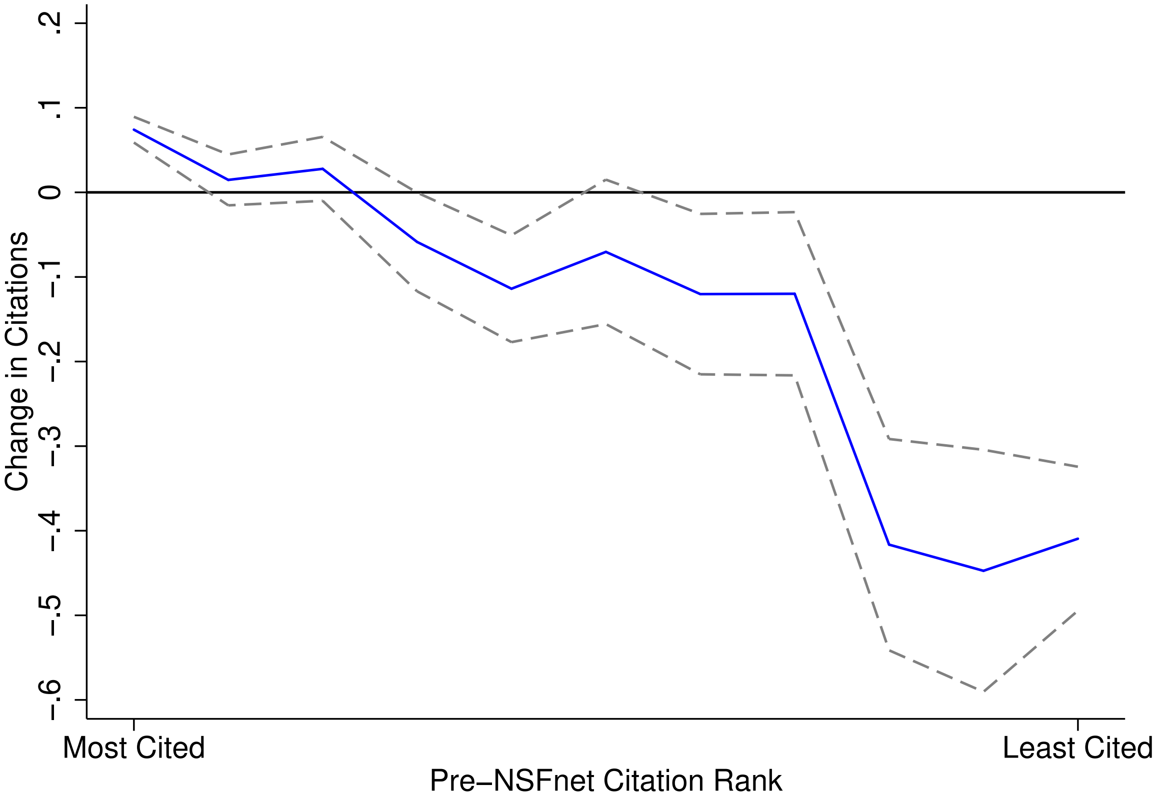
Rank	Field Cited	Citations
Most Cited	Physics	90,595
2.	Chemistry	3,121
3.	Astronomy	2,026
4.	Biology	1,144
5.	Medicine	698
6.	Math and Stat.	608
7.	Electrical Eng.	602
8.	Other Eng.	384
9.	Earth Sci.	322
Least Cited	All Others	

I bin all fields together ranked in tenth place or below as “least cited”. For most fields in the least cited bin, the total citations are zero or less than ten.

---

[29] For example, a ranking with two fields tied for second place each are placed at second.

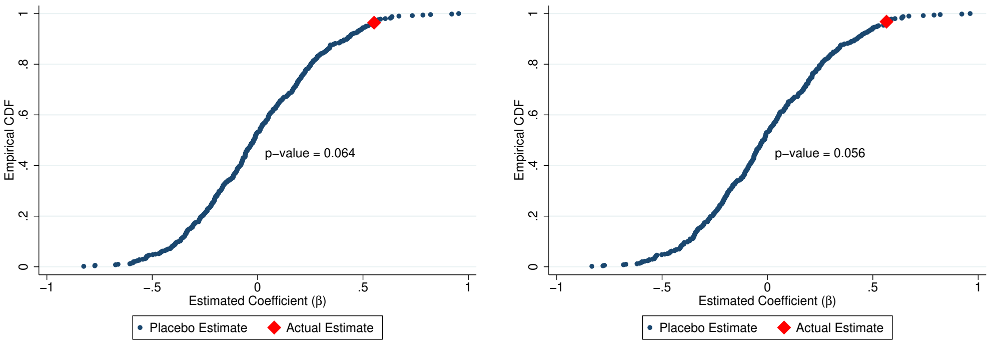
Figure A.1: Change in Citations by Field Citation Rank



Note: Figure shows the estimated change in citing another institution across differing scientific fields following institution pair's connection to the NSFnet. The figure plots estimates from a Poisson regression and 95% confidence intervals in blue and gray lines, respectively. The Pre-NSFnet citation rank orders the amount of citations made by an individual scientific field prior to the NSFnet program. The highest rank, or most cited field, is always within field (e.g. Physics citing Physics). The figure indicates that, following NSFnet connection, the amount of citations among the two most cited fields increases while the amount of citations among all other fields declines. Standard errors are clustered at the institution-pair level.

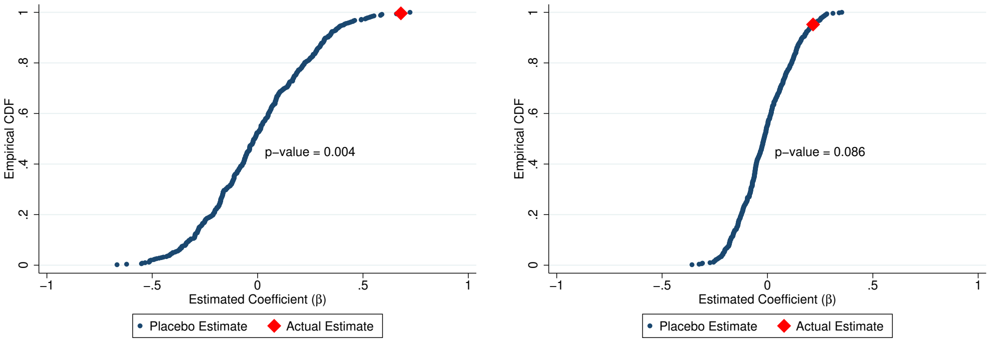
A.3 Placebo Test: Additional Estimates

Figure A.2: Distribution of Placebo Estimates



(a) Specification 1

(b) Specification 2



(c) Specification 3

(d) Specification 4