Quality of Communications Infrastructure Provision and Local Structural Transformation

Camilo Acosta [*], Luis Baldomero-Quintana [†‡]

September 16, 2022

PRELIMINARY WORK - PLEASE DO NOT CITE OR DISTRIBUTE

Abstract

We document that the quality of communication infrastructure has a causal impact in the economic structure of American counties, using Internet as example. We use as our instrumental variable the structure of ARPANET in 1979 (the precursor of the Internet backbone), a network funded by the Department of Defense. We determine ARPANET's location using historical government documents. Higher quality of internet leads to higher county GDP, employment, and higher shares of employment in high-skilled services, while negatively affecting activity in finance, retail, accommodation and food services. Two mechanisms can explain out results: input-output linkages and an increase in the usage of workers in ICT-intensive occupations.

JEL Codes:

Keywords: history of technology, communication technology, Internet, infrastructure, local structural transformation, spatial economics

[*] School of Finance, Economics, and Government, Universidad EAFIT, cacosta7@eafit.edu.co

[†] Department of Economics, William & Mary, labaldomeroqui@wm.edu

[‡] This project was funded by William & Mary Arts & Sciences Grant Funds 2021. We are thankful to Vint Cerf and Bob Kahn, the creators of the TCP/IP protocol, for sharing their knowledge about the ARPANET network. This paper benefited from conversations with Barthelemy Bonadio, Marco Gonzalez-Navarro, Xian Jiang, John Lopresti, Frédéric Robert-Nicoud, and William Strange. Eli Rothleder, Yiwen Sun, Maia Tindall, and Lauryn Walker provided extraordinary research assistance. All errors are our own.

You can see the computer age everywhere but in the productivity statistics
- Robert M. Solow

1 Introduction

Communication infrastructure differs from other types of infrastructure like roads, ports, or railroads. First, it increases the transmission speed of ideas between individuals as audiovisual content can be transferred in seconds. Second, it improves the technical functionality of computers, since multiple connected machines can solve difficult tasks faster. Third, it enhances the commercial uses of personal computers, since any individual with a digital device can consume services produced by other people. Thus, it is reasonable to expect that improvements in communication infrastructure induce local economic changes. These benefits have been acknowledged by U.S. lawmakers, who recently provided funds to improve communication infrastructure in the country (The White House, 2022).

In this paper, we document how differences in the quality of communication infrastructure influence the structural transformation of local economies. We provide causal evidence that higher quality of Internet generates local structural change in the U.S by favoring industries that are technology driven. To uncover this relationship, we use cross-sectional regressions at the county level in which the quality of Internet offered to business is our main treatment. For identification, we use as historical instrumental variable the distance to the lines connecting ARPANET nodes,[1] a network that was the backbone of Internet in its initial stage. These lines represent the actual telecommunications equipment installed to connect the network nodes. [2] Our estimates suggest that if a county improves its Internet quality, its GDP increases, as well as its total employment, average wages and its share of workers in high-skilled services.

For our empirical analysis, we combine historical data of the origins of Internet with several other sources. From the Census Bureau, we obtain county level economic outcomes such as GDP, total and sectoral employment and payroll. We also include data that measures the current quality of the provision of Internet to businesses for all U.S. counties from the Federal Communication Commission. Finally, we include other economic and geographic covariates from the 1970s and 1980s Census to control for historical factors that affect local economic activity.

Obtaining causal estimates of the impact of Internet quality provision on local outcomes is challenging. On one hand, counties with better amenities or higher productivity attract more college-

[1]In computer networks, a node is a connection point in a network that is a processing device with an assigned address, as a router, computer terminal, peripheral device, or mobile device, (Encyclopedia.com, 2022).

[2]ARPANET engineers did not display in their maps the location of the actual cables and equipment, only the nodes, only direct and indirect connections.

educated workers and productive establishments (Glaeser et al., 2004), thus Internet Service Providers (ISPs) are willing to offer better service in these locations. On the other hand, counties with low competition among ISPs may provide worst Internet service, a situation that disincentivizes establishments that need high-quality Internet to locate in these areas. We circumvent these issues by using as instrumental variable the spatial structure of the ARPANET, a network created by the Department of Defense (DoD) and the precursor of Internet. We use historical official maps and reports of the DoD office in charge of its design (Defense Advanced Research Projects Agency or DARPA, formerly known as ARPA) to find the exact location of the network nodes and their connections. We complement these documents by contacting former ARPANET researchers and Internet pioneers who participated in the creation of these maps.

Our instrument uses an idea by Forman et al. (2012) who use ARPANET nodes as an instrument for firms' investment on Internet to analyze how Internet impacts wages of U.S. counties in 1995-2000. Their IV strategy produced a weak instrument because nodes were scarce and private investment is lumpy. We adjust Forman et al. (2012) idea. We used the lines that connected the nodes as an IV, since they represent the high quality telecommunications equipment that ARPANET needed in the 1970s (Hauben et al., 1998). Moreover, our treatment is different too. We use as a treatment a smooth variable that measures quality of Internet provision rather than investment. These changes to the idea of Forman et al. (2012) produce a strong instrument that can be used for any cross-sectional analysis of how Internet impacts economic outcomes at the county level.[3]

Our instrument arguably satisfies the relevance, exogeneity, and the exclusion restriction assumptions. First, the instrument is relevant since literature on the history of Internet shows that from 1969 to the early 1980s, ARPANET physical infrastructure (represented by both the nodes and the lines) was the backbone of the Internet. Due to path dependence of infrastructure (Duranton et al., 2014; Duranton, 2015), the old Internet backbone is likely to be a good predictor of the modern Internet backbone, which we cannot observe.[4]. Given that physical closeness to the modern Internet backbone allows telecommunications companies to lower costs, we expect that locations closer to the former ARPANET lines are more likely to have to higher speed Internet. Second, the instrument satisfies the exclusion restriction given the historical context of the foundation of ARPANET. Official DoD reports suggest that DARPA decided the location of ARPANET nodes

[3] Jiang (2022) uses the original idea of Forman et al. (2012) in a DiD-IV setting to analyze firm outcomes. She also uses as an instrument the nodes of a network, NSFNET, which is considered the successor of ARPANET as an Internet backbone. Our historical context provides details on why for a cross-sectional analysis of county outcomes, a researcher must use the ARPANET network structure.

[4] There are two reasons for this. First, the information belongs to private companies. In current times, the Internet backbone equipment is built and maintained by private top tiers who provide connections to well-known commercial Internet service providers (Verizon, Comcast, T-Mobile, Xfinity, etc.). Second, the Internet backbone is considered of national security interest. For example, international affairs think tanks have lobbied to provide more safeguards to the transatlantic cables that are part of the worldwide Internet backbone. Such actions would protect the Internet (Wall, 2021).

based on computing research agendas, work philosophies of computer science departments, and DoD contracting relationships. Commercial companies did not influence the creation and design of ARPANET. Finally, we expect the instrument to satisfy the exogeneity assumption since current productivity and amenities of counties are unlikely to be related to whether ARPANET lines are onto these counties. This is because a straight line is the least cost way to connect two distant locations with infrastructure (roads, cables, railway tracks, tubes).

We find that higher quality of Internet provision leads to changes in the economic structure of U.S. counties. Specifically, our estimates indicate that higher Internet speeds has positive and statistically significant impacts on the share of local employment in five sectors: *Professional, Scientific, and Technical Services*; *Management of Companies and Enterprises*; *Educational Services*; *Administrative Services, Support Services, Waste Management and Remediation Services*; and *Other Services*. Interestingly, better provision of Internet has negative and statistically significant impacts on the employment share of *Finance and Insurance*; *Healthcare and Social Assistance*; *Retail Trade*; and *Accommodation and Food Services*. These effects are similar if we consider payroll shares instead of employment shares. Our results suggest telecommunication infrastructure favors the services sector. This is in line with the predictions of Beyers (2002) two decades ago on how ICT would generate a rise of new types of jobs in the services sector.

Higher Internet speed can have heterogeneous impacts within a broad sector. We use the case of Finance and Insurance (NAICS 52) to illustrate this. Although better provision of Internet reduces a county's share of employment of this sector, we find interesting patterns at a more granular level. On one side, better Internet reduces the county employment in *Credit Intermediation and Related Activities* (NAICS 523), which mostly includes workers in physical banks. This result seems intuitive given that online banking and local bank branches are imperfect substitutes, and most locations, even small towns, have a physical bank. On the other side, higher Internet speed increases local employment in *Securities, Commodity Contracts, Other Financial Investments and Related Activities* (NAICS 522), which includes technology-intensive financial services. Since Internet reduces the costs of financial advisors across the U.S. of acquiring information, better ICT allows them to expand.

We proposed two mechanisms to explain these effects: input-output linkages and rise in the employment of workers who use relatively more ICT. Our findings show that sectors that are highly dependent of ICT-inputs drive the county-level results, while the impacts are negative or not statistically significant for sectors with little dependence on ICT-inputs. Second, we provide evidence that better Internet provision in a county increases the number of workers that work in occupations that might benefit from the use of ICT technology, such as management, business and finance, office and administrative support, engineering, and computer sciences. Notice that the sectors that benefit the most from higher quality of Internet provision (business and administrative services, education and professional services) benefit from higher abundance of workers who are in related occupations (managers, business, finance and administrative support, and computer science).

Our results have clear implications for infrastructure policy and public expenditures. Specifically, they suggest that Internet quality can explain regional inequality patterns. Advanced nations have used public funds to improve the quality of communications infrastructure via the promotion of universal broadband policies, including Canada (Government of Canada, 2019), the United States (The White House, 2022), Germany (European Commission, 2022) and the UK (Hutton, 2022). The motivation behind these policies is to reduce the technology education gap. We provide evidence that public plans that improve the quality of communications infrastructure across regions can also address concerns regarding regional inequality, even in advanced nations.

We contribute to the literature that studies the economic impacts of Internet and communication technologies on national growth, regional economies, trade flows, and firm-level outcomes. Our main idea relates to previous work analyzing how ICT impact growth using cross-sectional country-level data (Greenstein and Spiller, 1996; Hardy, 1980; Norton, 1992; Röller and Waverman, 2001; Oestmann and Bennöhr, 2015). We complement this work by providing causal estimates of the effect of communication technologies local economic structure, and by highlighting the importance of quality of the infrastructure. Lastly, we focus on the effects of quality of Internet (extensive margin), while the majority of the existing work focuses on availability of Internet (intensive margin). Our results provide evidence that the extensive margin of the provision of Internet is economically relevant.

Existing work on economic geography and urban economics has focused on how communication costs, ICT, and the diffusion of ideas shape agglomeration patterns and determines city structure. This includes work on the ICT impact on cities (Gaspar and Glaeser, 1998; Glaeser and Ponzetto, 2007); the spatial expansion of Internet across cities (Malecki, 2002); how Internet affects property prices (Ahlfeldt et al., 7 01; Ford et al., 2005a; Dietzel, 2016; Beracha and Wintoki, 2013); how Internet startups agglomerated in specific cities even when the Internet reduced communication costs and became a dispersion forces (Zook, 2002); the role of a better transmission of ideas on innovation (Carlino et al., 2007; Rosenblat and Mobius, 2004); how new jobs created by the ICT revolution are subject to agglomeration forces (Lin, 5 01); and how transportation and communication costs reduce spatial frictions for the diffusion of knowledge that can improve the productivity of isolated farms in the U.S. (Kantor and Whalley, 2019). Interestingly, Charlot and Duranton (2006) find that causality can go in the other direction: cities can foster the use of telecommunication technologies. We contribute to this literature by showing how cheaper communication and the quality of internet are key to understand local structural transformation.

Recent work has also explored how communication technology impacts firms. In particular, this literature includes papers studying how broadband access positively affects entrepreneurship (Beem, 2022), R&D collaboration within firms (Forman and Van Zeebroeck, 2012), and firm size in both urban and rural areas (DeStefano et al., 2 04). Other work includes how digital platforms benefit small and medium enterprises (Arfi and Hikkerova, 2021). Firm level improvements coming from reductions in communication costs, are not only due to internet expansion, but also due to the expansion of other technologies. In particular, Marinoni and Roche (2022) document that the expansion of the postal service in 1880-1990 contributed significantly to firm creation and firm performance. Moreover, evidence shows that communication costs are an important determinant of the spatial organization of multi-establishment firms (Acosta and Lyngemark, 2020; Jiang, 2022). We contribute to this literature by providing evidence that within the same region, communication technology changes economic structure. Our results suggest that the impact of communication costs on firms might depend on the sector of the firm.

Existing work on international economics studies how trade is influenced by overall communication costs (Fink et al., 2005; Allen, 2014), and Internet (Freund and Weinhold, 2004; Blum and Goldfarb, 2006,?; Breinlich and Criscuolo, 2011). Existing papers on the intersection of economic history and international trade show that historical changes in communication technology led to higher trade flows (Juhász and Steinwender, 2018; Steinwender, 2018), and changed the comparative advantage of regions (Eckert et al., 2020). Other work documents that lower travel costs facilitate face-to-face communication, which itself impacts trade flows (Cristea, 2011). Our paper complements previous work in two ways. First, we provide evidence that the quality of communications infrastructure matters, and not only access to it. Second we analyze how communication costs impact structural transformation and economic activity as a whole, not only trade flows or trade costs.

Our work has similarities with two papers: Forman et al. (2012) and Jiang (2022). Forman et al. (2012) analyze how private investment on Internet by large firms increased wages only in high-income U.S. counties between 1995 and 2000. We complement their work by focusing on how Internet provision changes the local structure of the economy in modern times. Differently, we find that better Internet provision to firms increases average GDP and wages. Our differences arise from the type of treatment, instrumental variable used and the period of analysis. While we study the impacts of Internet in 2018, more than 20 years after the initial Internet boom in 1995, Forman et al. (2012) focuses on short-run effects. Unlike the 1990s, as of 2018 Internet was a consolidated technology with high levels of adoption. Furthermore, since 2000 Internet has produced additional technological and economic changes.[5]

[5] First, speed increased exponentially. Connections used in the 1990s were based on the most common technology (dial-up) with download and upload speeds of at most 56 kpbs. The average speeds in American counties in 2018 are between 2 and 671.1 Mbphs for download, and between 1.1 and 615.7 mpbs for upload. If we consider the

Lastly, Forman et al. (2012) consider as their main IV the number of programmers in other counties in the same firm, and for a robustness check they use as a second IV the location of ARPANET nodes. We focus on the lines connecting the nodes for two reasons, First, when Forman et al. (2012) uses nodes as IV for private Internet investment, he obtains a weak instrument. But when we use ARPANET lines instead of nodes, we obtain a strong instrument. The use of lines as IV is consistent with the significant investments of DARPA to connect the nodes with high quality telecommunications equipment (DARPA, 1981; McKenzie and Walden, 1991). Lastly, another contribution is that we complement the weak instrument of Forman et al. (2012), by creating an instrument that is stronger and that can be broadly applied by empirical researchers.

Our work is related to the empirical framework Jiang (2022), although our research questions differ. She analyzes how Internet popularity changed the spatial organization of manufacturing firms in the 1990s. She uses an IV-DiD approach to motivate a structural model. Her spatial IV is the distance to NSFNET nodes in 1994 (the network that overtook the role of Internet backbone after ARPANET was decommissioned). Her instrument is based on Forman et al. (2012). We chose ARPANET instead of NSFNET because historical sources show that commercial interests influenced NSFNET regional networks. Hence, the location of NSFNET nodes is plausibly correlated with counties' productivity in the 1990s and today. This is not the case of ARPANET, since the network was only subject to national defense interests Abbate (2000); DARPA (1981); Hauben et al. (1998); Leiner et al. (1997). Even we cannot use NSFNET nodes, they are adequate for Jiang (2022) empirical strategy. [6]

The rest of the paper proceeds as follows. In Section 2, we document the historical origins of ARPANET, which is the precursor of the Internet backbone. In Section 3, we describe our data and present descriptive statistics. Section 4 presents our empirical model, including our identification strategy. In Section 5 we present the results of the paper, together with the two mechanisms proposed that can explain such results. Section 6 concludes.

lower bounds, speeds are 17 to 35 times faster in 2018 relative to dial-up technology. Moreover, social media and better search engines changed markets (Dietzel, 2016; Beracha and Wintoki, 2013; Ford et al., 2005b; Oestmann and Bennöhr, 2015), the rise of Amazon transformed the retail sector (Cavallo, 2018), and video and mobile devices modified consumption and working patterns (Bresnahan et al., 2014; Dingel and Neiman, 2020).

[6] Correlation of NSFNET nodes with county productivity does not represent an identification issue for Jiang (2022). First, her regressions use outcomes for individual manufacturing firms, whose location determinants are probably unrelated to ICT technologies before the mid-1990s. Second, her framework requires changes between two spatial equilibria in the internal organization of firms. Her empirical framework satisfies these assumptions. If researchers were to use her identification strategy to analyze the internal organization of service firms or county level outcomes, some adjustments to her empirical framework would be required.

2 Historical Context: ARPANET

In this section we document the historical origins of the Internet. We provide historical information on ARPANET, the precursor of Internet. We also present facts about the successor of ARPANET, a National Sciences Foundation network called NSFNET. We focus on the factors that influenced the spatial configuration of both ARPANET and NSFNET. This chronicle supports the idea that the location of ARPANET nodes are exogenous to local productivity while NSFNET nodes are not.

Origin of ARPANET. ARPANET was a research project financed and developed by DARPA, a research office of the DoD. DARPA has as objective to create a network that connected computing centers. Two factors drove the financing of the project. First, the DoD had a strong interest in financing research after the Soviet Union launched its first satellite. Second, the DoD wanted its computers to share data to lower administrative costs. In the early 1960s, the DoD had thousands of computers that operated autonomously without any connection, which generated large costs since all data files and software had to be reprogrammed in every computer, and military personnel had to be trained to use devices from different manufacturers. All these costs were more than double the national budget for software creation and maintenance, (DARPA, 1981).

DARPA invested resources for computing network research after hiring a prominent innovator in computer science, Dr. J.C.R. Licklider. He was one of the first scientists to propose the idea of using computers to improve human communication (Licklider and Taylor, 1968). In the 1960s, computing companies and most U.S. academic computer science departments focused on improving the speed of computers to complete tasks, a concept known as batch processing (Hardy, 1980). When Licklider arrived at DARPA to design a computer network, he shifted all computer science research contracts away from private companies to academic departments with interests in network research. He considered that private sector interests were focused exclusively on batch processing. In this way, some computer research centers at U.S. universities became DoD contractors via DARPA. Commercial interests play no role in the initial design and development of ARPANET (DARPA, 1981,?).

The main idea behind ARPANET and the Internet was to connect independent local networks of arbitrary and different designs (DARPA, 1981; Leiner et al., 1997). Thus, if one local network was lost the whole system continued working. Establishing a network like ARPANET demanded novel technological resources: a specific software to define a protocol (set of rules for data transmission) and equipment to guarantee the stability and delay of a network (the time for a signal to traverse a network). During the late 1960s, prior to the first ARPANET connection in 1969, DARPA researchers focused on solving these issues.[7]

DARPA initially considered 19 possible participants to be included in the first connection (Hauben et al., 1998). The first four nodes connected in 1969 with research centers that were DoD contractors. The nodes were chosen due to specific technical expertise. [8] The first nodes were connected in 1969, and the network was fully operational in 1971. Figure 1 presents the original description of the network as described by Vint Cerf and Bob Kahn (two ARPANET researchers and Internet pioneers) in December 1969.

Throughout the 1970s, maintaining ARPANET required require (i) financial resources by the DoD and (ii) frontier technological expertise, (DARPA, 1981; Hauben et al., 1998; Leiner et al., 1997). Academic researchers had to solve novel engineering issues frequently. Researchers working on ARPANET collaborated even if they belong to different universities and they created open documentation of the new technologies,(Leiner et al., 1997). [9] Thus, ARPANET nodes also depended on the work style of computer science departments, since an uncooperative computer science department was less likely to become an ARPANET node.

In conclusion, three factors influenced the location of ARPANET nodes. First, in the 1970s only universities with DoD contracts were connected to ARPANET, (Abbate, 2000). Second, universities with a research agenda focused on computer networks were more likely to participate in ARPANET, since research on batch processing dominated computer science across U.S. universities, (Hauben et al., 1998). Third, only universities willing to collaborate with competitors and to share publicly research findings would join the project. Thus, we can conclude that the location of ARPANET nodes in the 1970s was driven by factors unique to national defense and computer science research, and commercial interests had zero influence on the network (DARPA, 1981).

From ARPANET to NSFNET. In 1980, only nodes related to defense computing research or operations were connected to the Internet. The situation was different in the late 1980s: the backbone was transferred from military to civilian control, and Internet users included universities, commercial interests and federal agencies (Abbate, 2000). Five events influenced this change. First, DoD created a separate military computing network in 1983, MILNET. Afterward, ARPANET only had

[7] In an experiment in 1966-1967, an MIT laboratory and military contractor SDC in Santa Monica, CA, connected two computers of different manufacturers using the existing telephone circuit technology. This experiment showed that existing telephone circuits could not provide communication between devices of good quality (Hauben et al., 1998; Leiner et al., 1997).

[8] These four nodes correspond to the University of California—Los Angeles, University of California–Santa Barbara, Stanford Research Institute, and the University of Utah.

[9] Leiner et al. (1997) considers that ARPANET's technical notes about engineering network design, which were open to the public, called "Request for Comments", were a driving force of the network's success. The notes were open to the public even if the DoD financed the network (Leiner et al., 1997; Hauben et al., 1998). Collaboration and open documentation allowed faster feedback between ARPANET researchers which lead to quick solutions of engineering issues. New findings were available to the public.

research centers as nodes. Second, NSF funded NSFNET, a national network of regional academic networks. Third, DARPA and NSF collaborated to guarantee interoperability between ARPANET and NSFNET. Fourth, NSF encouraged NSFNET regional networks to provide service to commercial companies. Fifth, ARPANET transferred the role as backbone of Internet to NSFNET in 1988, and ARPANET was decommissioned in 1990 (Abbate, 2000; Leiner et al., 1997).

During the 1980s, ARPANET opened the network to more academic centers since DARPA desired to transfer of network technology to the entire academic community. The agency also wanted to transfer at some point the backbone of the network[10] to another group, and to promote the adoption of this network by other government agencies, (DARPA, 1981). DARPA eventually allowed local university networks to connect to ARPANET. These local networks consisted of computers connected within research institutes or across a group of universities sharing computing resources.

In 1981, a consortium of universities that were not DoD contractors requested a grant from the NSF to create CSNET, an academic network implemented in the period 1981-1984. In 1981, DARPA signed a cooperative agreement with CSNET to share ARPANET's infrastructure (McKenzie and Walden, 1991; Leiner et al., 1997). To connect to CSNET, NSF agreed to provide resources, but each university or research center would independently manage their internal network of computers. CSNET was possible due to NSF funding and the existence of universities that shared computer resources since the 1970s. CSNET is considered the predecessor of NSFNET.

In 1985 NSF starts building NSFNET, a national network that connected regional networks created by universities[11]. The objective of NSFNET was to create a two-tier system: NSF-funded regional networks, and a national backbone network connected them. NSF outsourced a contract to build the physical infrastructure in 1985, and its construction took three years (Abbate, 2000). In 1986, NSF and DARPA created a technical document to guarantee that ARPANET and the regional NSF networks had interoperability (Leiner et al., 1997). Later, in 1987, the NSF reached an agreement with ARPANET to use ARPANET as a national backbone to connect regional networks, in exchange for sharing ARPANET's operating costs (Abbate, 2000). The NSFNET national backbone started functions in 1988.

Commercial interests and spatial location of NSFNET. While its initial years ARPANET was only available for research centers and a couple of military contractors, without commercial interests interfering with it, this was not the case for NSFNET due to the incentives imposed

[10] The Internet is de facto a network of local networks. It is divided in three tiers. The upper tier is considered a backbone because it reaches all local networks.

[11] The regional networks were BARRNet for the San Francisco Area, MIDNet in the Midwest, WESNET in the Rocky Mountain region, USAN for those research centers connected to the National Center for Atmospheric Research, NorthWestNet in the Northwest, NYSERNet in the New York state area, Sesquinet in Texas, SURANet in the Southeast. NSFNET also connected large computer centers in San Diego, University of Illinois, and Pittsburgh.

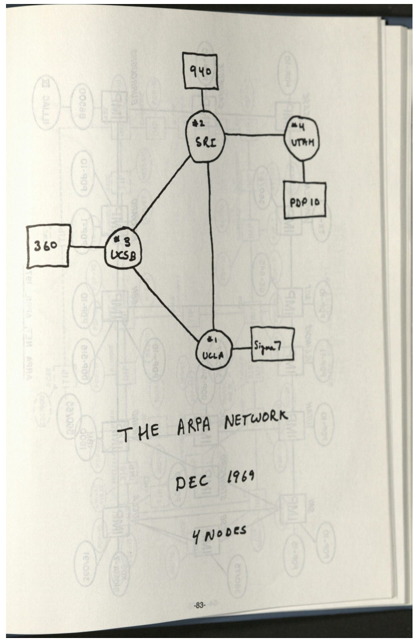
during its creation. Although Federal law prohibited commercial companies to use the NSF national backbone (U.S. Congress, 1992), private companies could connect to regional networks (Leiner et al., 1997; McKenzie and Walden, 1991). Furthermore, NSF funded regional networks with the condition that they would be financially autonomous in three years (McKenzie and Walden, 1991). NSF also encouraged regional networks to find commercial clients(Leiner et al., 1997). Abbate (2000) provides evidence that the NSF-funding of regional networks created private Internet providers in the 1980s, way before the privatization of NSFNET in 1995.

The NSF wanted private companies to build national Internet networks (Leiner et al., 1997). Although commercial networks emerged, no national private backbone did until the privatization of NSFNET in the early 1990s. The NSF financial incentives created a situation in which groups of universities would request funding to create a regional network only if they had increasing returns to scale. Hence, universities would request funds only if they were large enough or if they could get commercial clients from nearby locations. Therefore, NSFNET nodes are likely to be spatially correlated with the productivity of cities, counties or regions, unlike ARPANET nodes.

End of ARPANET, NSFNET as the Internet backbone, and the privatization of NSFNET. In the late 1980s, the exponential growth of Internet users revealed the age and obsolescence of ARPANET as the Internet backbone, and DARPA decided to decommission the program. The creation of a new backbone network was considered as a replacement (McKenzie and Walden, 1991; Abbate, 2000), but the final resolution was simpler: the NSFNET national network had enough capacity to overtake the role of ARPANET as the internet backbone. The protocol designed by Vint Cerf and Robert Kahn for ARPANET made the transition smooth (TCP/IP). Between 1988 and 1989, DARPA sites transferred their host connections from ARPANET to NSFNET, and in February 1990 ARPANET was formally decommissioned (Abbate, 2000). Independent commercial networks began to grow in the late 1980s and early 1990s due to the NSFNET regional network' services provided to commercial interests (Abbate, 2000). In 1995, the NSF was scheduled to transfer the NSF national backbone to private companies. Harris and Gerich (1996) documents the technical details of this transition.

The history of Internet allows us to reach three conclusions related to our instrumental variable. First, ARPANET was not influence by commercial interests. Its existence and design was driven by the resources invested by the DoD. Second, the location of nodes was driven by DoD contracts, and the research agenda and work philosophies of computer science departments in the 1960s and 1970s. Third, NSFNET regional networks were influenced by commercial interests even if private companies could not use the NSFNET national backbone infrastructure. Thus, the NSFNET nodes are likely to be correlated with local productivity of counties.

Figure 1: Original description of ARPANET nodes in December 1969



Source: Cerf and Khan (1990)

3 Data

Our data comes from four different sources: the US Census Bureau, the Bureau of Economic Analysis, the Federal Communications Commission, and decommissioned maps created by the Department of Defense (DoD). In this subsection we describe in more detail each of these data sources. Afterward, we show some descriptive statistics for the main variables of interest.

First, we use employment counts and aggregate payroll by county and sector from the US Census Bureau's 2018 County Business Patterns (CBP). The CBP includes the number of establishments and employment for every county during the second week of March, as well as first quarter and annual payroll. Using these variables, we also compute the average wage within counties as the ratio between annual payroll and total employment. The CBP data also contain information by NAICS (North America Industry Classification System) sector, up to 6-digits. However, due to confidentiality restrictions, a data point is only published if it contains at least three establishments. Thus, to minimize the number of missing values arising from these restrictions, we use sectoral data at the 2- and 3-digit NAICS code.[12] For each sector, we compute the share of employment and payroll in each sector within each county. Finally, we complement these data with information on each county's GDP in 2018 from the Bureau of Economic Analysis.

[12]The full list of 2-digit NAICS sector is included in Table A-1 in the Appendix-

Second, to capture today's access and quality of the Internet, we use data from the Federal Communications Commission (FCC) for 2018. These data comes from FCC's Form 477 which has to be filled by all broadband providers twice a year (Federal Communications Commission, 2019). In these data, fixed providers report the list of census blocks in which they offer a particular service to at least one location. Therefore, the data contain for each census block inside the US, a list of all internet providers, together with the offered technology of transmission (e.g., cable model, xDSL, fiber, fixed wireless, etc.), the maximum advertised download and upload speeds/bandwidth (for consumers) and the maximum download and upload speeds/bandwidth (for businesses) for each technology.[13] From these data, we keep only those providers and technologies that were offered to businesses, with which we compute the average and the median download and upload speeds for each county.

Third, to build our instrumental variable, we digitized images of decommissioned DoD documents that contained ARPANET maps. In 1990, Vint Cerf and Robert Kahn (ARPANET pioneers) collected these maps and published them in the Journal of the Association for Computing Machinery (Cerf and Khan, 1990), which are only available in the physical version of the journal. These maps included the abbreviated name and the location of the nodes, and the lines connecting them, which represent the network connections between ARPANET nodes. As the name of the nodes is not available in the journal (only their abbreviation), we proceeded to analyze DoD decommissioned reports and other historical sources. We also contacted Vint Cerf and Bob Kahn who share their knowledge about the ARPANET network, including the names of some remaining nodes. Figure 2 shows the original and the digitized maps for 1979. Using the information contained in these maps, we compute the minimum distance between each county's centroid and one of the connection lines, together with an indicator variable that equals 1 if a county contains a node (and 0 otherwise), and another that equals 1 if it contains a connection line.

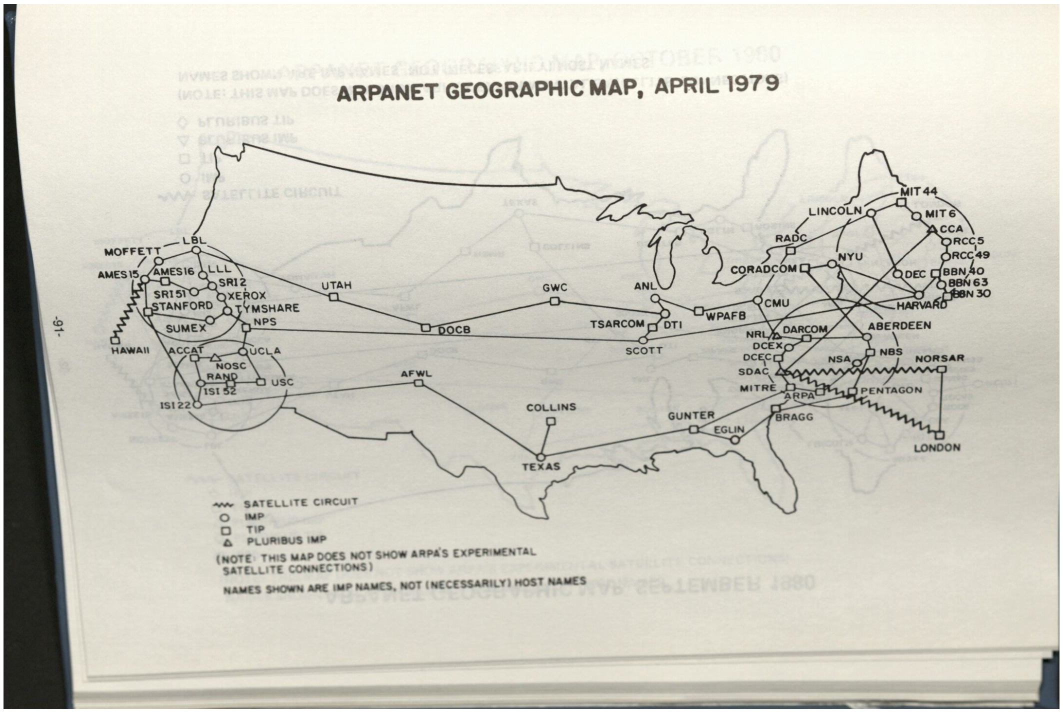
Finally, we also include different controls for population and industrial composition in 1980 from the US Census. In particular, we computed for the 1970's and 1980's share of workers in agriculture and mining, manufacturing, high-skilled services (i.e., information, finance, insurance, real estate, professional, management, education and health) and other services (i.e., utilities, construction, retail, wholesale, transportation, support, entertainment, accommodation and food services). Moreover, we create a series of dummy variables that indicate whether a county lies in the border with Canada or Mexico, or along a coast (either oceanic or one of the Great Lakes).

In Table 1, we present the average and median values for some of the variables of interest. Re-

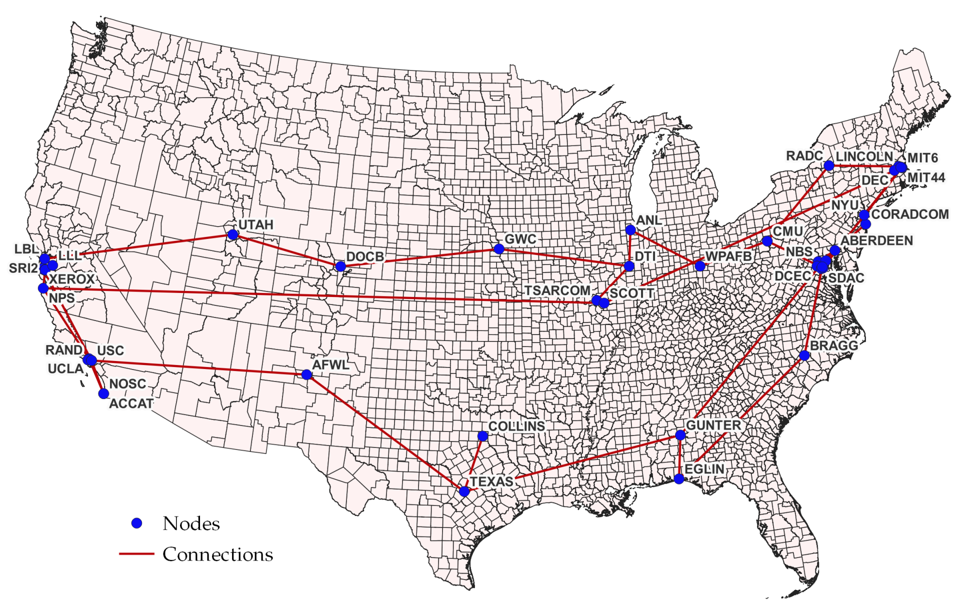
[13] As the name suggests, download speed corresponds to the speed used to download data from a server to a computer in the form of text, files, audio, images, videos, etc. On the other hand, upload speed refers to how fast can information be sent (in the form of text, audio, images, etc.) from a computer to another device connected to the internet.

Figure 2: ARPANET - 1979

Panel A: Original Map



Panel B: Digitized Map



Note: this figure shows

garding current economic activity, notice that around half of employment and total payroll in the average county is generated in the "other services" category, which includes wholesale and retail trade, administration and support, construction, accommodation and food services, among others. These services are followed by "high-skilled services", which include information, management, professional, educational services, among others, generating a third of the economic activity. most of the counties' employment. Manufacturing lies in a third place with around 15% of the activity. For all variables, the mean and the median are quite similar. Figures ?? and ?? show the spatial distribution of these shares across the US. Regarding internet, the average download and upload speeds are 97 and 86 megabits per second (Mbps), respectively, while in the median county they are 39 and 28 Mbps, respectively. Finally, in 1979 only 1% of counties had an ARPANET node, while 14% of them had a line going through; and the average county was 192km away from a line.

Table 1: Descriptive Statistics

Category	**Average across Counties**	**Mean**	**Median**
Current Economic Activity	Share agriculture & mining employees	0.02	0.00
	Share manufacturing employees	0.15	0.12
	Share high-skilled services employees	0.30	0.29
	Share other services employees	0.53	0.52
	Share agriculture & mining payroll	0.04	0.00
	Share manufacturing payroll	0.19	0.16
	Share high-skilled services payroll	0.33	0.32
	Share other services payroll	0.44	0.42
	Total GDP (millions)	6,507	1,021
	Total Employment	39,826	6,557
	Total Payroll (millions)	2,163	248
	Average Wage (thousands)	39.71	37.83
Current Internet	Mean download speed (Mbps)	96.50	38.87
	Mean upload speed (Mbps)	85.97	27.86
ARPANET	Distance to APANET lines in 1979 (km)	191.78	119.74
	Had node in 1979	0.01	0.00
	Had connection line in 1979	0.14	0.00

Notes: This table shows the mean and the median of our variables of interest averaging across all counties in continental US. Variables regarding current economic activity and internet speeds correspond to 2018. Speeds are measured in Megabits per second (Mbps). High-skilled services include information; finance and insurance; real estate services; professional, scientific and technical services; management of companies and enterprises; educational services; and health care and social assistance. Other services include utilities; construction; wholesale trade; retail trade; transportation and warehousing; administrative and support and waste management and remediation services; arts, entertainment and recreation; accommodation and food services; and other services. ...

4 Empirical Model and Identification

In this section we discuss the empirical framework we use to estimate the effects of quality of Internet provision on current local economic structure. Moreover, we weigh on the use of ARPANET lines as an instrumental variable, and we asses the assumptions we need to interpret our results as causal.

To measure the effects of Internet quality on modern county economic outcomes, we estimate the following model:

$$Y_c = \alpha + \beta log(InternetQuality_c) + X_c'\Theta + \epsilon_c \tag{1}$$

where Y_c is a local economic outcome in county c in 2018, such as the log of total GDP or the share of employment in services subsectors; $InternetQuality_c$ is a measure of Internet speed (mean or median, upload or download) in county c in 2018; and X_c' is a group of geographical and economic characteristics from the 1980s (pre-treatment); $epsilon_c$ considers the unobserved factors that impact local economic outcomes, such as productivity in county c.[14] The group of control variables includes: dummy variables that equal one if the county is next to the Canadian or Mexican border, and dummy variables that equal one if the county is in the coast of an ocean or a Great Lake. We also consider the total population and the local economic structure from 1980. In particular, we include the share of workers in agriculture, manufacturing, construction, high-skilled services and other services in that year. Our main parameter of interest is β, which identifies the elasticity (or semi-elasticity) of a local economic outcome with respect to the quality of Internet provision.

The estimation of equation (1) would not yield causal estimates as the relation between the variables of interest suffers from endogeneity. For instance, counties with higher productivity or better amenities (e.g., cultural, environmental, urban) might have a larger share of GDP coming from high-skilled services sectors since these counties can attract more productive service establishments or more high-skilled workers. Hence, the willingness to pay for better Internet is higher in these counties, and companies will provide better Internet quality. Thus, we are in the presence of upward bias. On the other hand, counties with better environmental aesthetic amenities (e.g., rivers, mountains, lakes, strath) might be able to attract more high-skilled workers which would generate incentives to ISPs to provide high speed Internet. But at the same time such amenities could make it more difficult to build physical telecommunications infrastructure thus lowering the quality of Internet provision in the county. In this case we are in the presence of downward bias.

To recover the causal parameter, we follow an instrumental variable approach using the spatial

[14] In all the estimations we use Spatial Heteroskedasticity and Autocorrelation Consistent (SHAC) standard errors as proposed by (Conley, 1999) with a ratio of 28.55km, which correspond to the ratio of a circle that would cover the surface of the median metropolitan statistical area in the U.S.

structure of ARPANET in 1979 (years before the Internet had commercial viability) as the source of exogenous variation.[15] As we document in Section 2, the initial nodes of ARPANET were influenced by whether the researchers in academic institutions were contractors for the Department of Defense (DoD) through its ARPA agency, the research agenda of the institution, and the collaborative work style of the researchers. Hence, the decisions regarding the location of the nodes, and thus the spatial structure of ARPANET, were based on research agendas, work styles, and scientific contractor processes with computer science academics hired by the DoD via their research office DARPA in the 1960s. DARPA saw computing research differently relative to the private sector and it prioritized network research, a novel concept at the time, as we discuss in section 2. Therefore, it is unlikely that the nodes of ARPANET were selected only in counties with the highest productivity levels. Moreover, we selected the status of the network in 1979, when ARPANET only connected DoD contractors and it was still under military management (Abbate, 2000). This guarantees that the structure of ARPANET was more closely related to the research interests of the DoD.

Exploiting the spatial structure of ARPANET, our first stage is defined by

$$log(InternetQuality_c) = a + b \cdot ARPANET_c + X_c'\Gamma + \nu_c \tag{2}$$

where $ARPANET_c$ denotes the log of the minimum distance between county's c centroid and an ARPANET connection line, which were presented in Figure 2. The connections line are proxies for the equipment that connected ARPANET nodes. Since government reports and maps do not contain any information regarding the physical infrastructure used to connect the nodes, a straight line that mimics the maps of ARPANET is our proxy for such physical infrastructure. Given that ARPANET had specific network requirements (reliability and delay) to guarantee interconnection quality between ARPANET computers, as we document in Section 2, the physical infrastructure that connected the nodes must have been of the highest quality for the 1970s. We estimate different specifications of equation (2) using other instruments, including a dummy variable that equals 1 if the lines that connect the ARPANET nodes cross county c, and distance categories to the closest ARPANET connection line.

Our identification strategy has a similar logic as Duranton et al. (2014) and Duranton (2015). They use historical routes in the U.S. and Colombia to instrument for the location of modern highways. As it was easier to build highways following historical paths, nowadays it is easier to provide better Internet in locations closest to the ARPANET connecting lines as ARPANET was at some point in the past the Internet backbone, (Abbate, 2000; Leiner et al., 1997; McKenzie and Walden, 1991). As we mention in the introduction, our strategy is similar to Forman et al. (2012) who consider private investment on Internet as treatment and use the ARPANET nodes as an instrument (Jiang

[15] The first private Internet service provider, The World, appeared in the 1989.

(2022) also uses a similar approach, but using NSFNET nodes in a DiD framework). We adjust the idea of Forman et al. (2012) because it produces a weak instrument since investment is lumpy and the nodes are scarce. We use the connection between the nodes, and we consider as our treatment the reported Internet speed of companies to the FCC. This produces a strong instrument. We do not consider NSFNET nodes, since historical evidence suggest they might be endogenous in a cross-sectional empirical framework that considers regions or counties as the unit of observation. [16] Therefore, we use the ARPANET historical documentation maps from 1979 for our empirical strategy, which contain all the connections between the nodes.

We consider that our instrumental variable is relevant since ARPANET is the first precursor of the Internet. In 1989, the NSF allowed the network to be used for commercial reasons. Although, the last-mile connections used the telephone network (also known as Tier 3), the main backbone of the commercial Internet relied on ARPANET nodes and lines in its early stages (Leiner et al., 1997). Hence, closeness to the Internet backbone today allows ISPs to provide cheaper high-quality Internet service. Since we cannot observe the Internet backbone today given that it is private information and due to security reasons, we use ARPANET lines as a predictor of the modern Internet backbone. Moreover, because it is cheaper to expand the backbone of the Internet based on the past ARPANET equipment due to costs reasons, we expect that for counties closer to existing ARPANET infrastructure in 1979 it became cheaper to provide early private access to the Internet in the 1990s, which has lead to better access until today via a path dependence mechanism.

We explore the positive correlation between the location of the ARPANET network and today's internet visually in Figure 3. These maps show that if a county is connected to ARPANET lines it has better Internet quality nowadays. Our results apply for both upload and download speeds. We formally test the relevance of our instrument estimating different specifications of equation (2) using OLS. In particular, Table 2 validates the previous results, even when we include the vector of geographic and 1980 economic characteristics. In particular, those counties that had an ARPANET node in 1979 have better internet today (Panel A); however, since less than 1% of counties had a node, using them as an IV might cause a weak instrument problem. In a similar way, counties that had an ARPANET connection line in 1979 have today download and upload speeds that are 27% and 38% higher, respectively, relative to counties without a line (Panel B). Noticeably, counties farther away from a 1979 ARPANET line have lower Internet speeds (Panel C). In particular, a 10% increase in distance is correlated with speeds between 1.5% and 2% lower. Lastly, this negative relationship with distance holds if we consider distance categories instead of a continuous measure

[16] NSF encouraged local NSFNET regional networks to provide service to private companies. Due to increasing returns to scale only some universities created NSFNET regional networks. Most likely, the ones where there was abundance of private companies that needed an advanced telecommunications service in the 1980s. This represent an issue for cross-sectional regressions at the county level. Nonetheless, NSFNET nodes can be used as IV for firm level outcomes like the ones of Jiang (2022).

(Panel D). These results support the relevance of our instrumental variable.

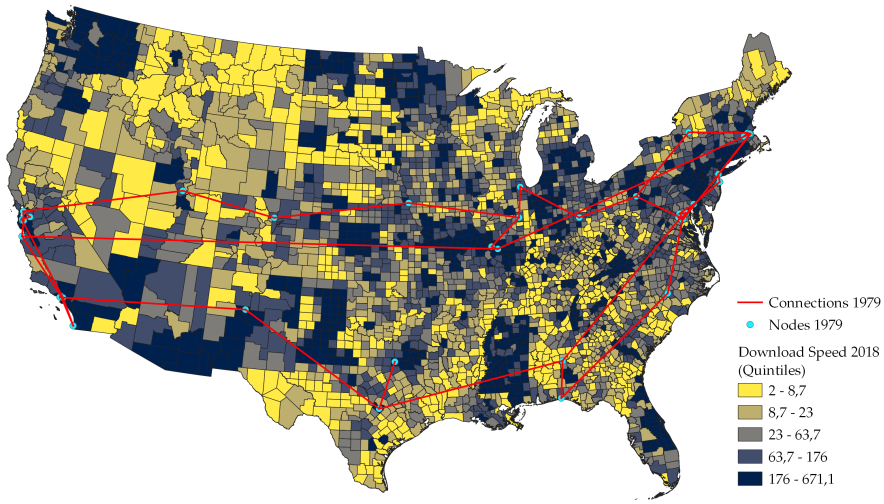
Table 2: The Impact of ARPANET on Current Internet

***Panel A**: County has a node*	Mean download speed		Mean upload speed	
	(1)	(2)	(3)	(4)
Had node in 1979	1.313***	0.458**	1.605***	0.591***
	(0.154)	(0.185)	(0.166)	(0.210)
Constant	3.599***	3.506***	3.237***	3.107***
	(0.039)	(0.049)	(0.046)	(0.059)
***Panel B**: County has a line*	Mean download speed		Mean upload speed	
	(1)	(2)	(3)	(4)
Had connection line in 1979	0.428***	0.267***	0.556***	0.375***
	(0.077)	(0.077)	(0.088)	(0.088)
Constant	3.552***	3.471***	3.174***	3.058***
	(0.041)	(0.049)	(0.048)	(0.059)
***Panel C**: Distance to a line*	Mean download speed		Mean upload speed	
	(1)	(2)	(3)	(4)
Distance to Line in 1979	-0.180***	-0.154***	-0.226***	-0.201***
	(0.025)	(0.025)	(0.028)	(0.029)
Constant	5.683***	5.262***	5.858***	5.409***
	(0.285)	(0.295)	(0.327)	(0.340)
***Panel D**: Distance to a line*	Mean download speed		Mean upload speed	
(Categories)	(1)	(2)	(3)	(4)
Has a line	0.715***	0.568***	0.903***	0.750***
	(0.106)	(0.108)	(0.120)	(0.122)
No line and $dist \in (0km, 73.4km]$	0.606***	0.610***	0.735***	0.763***
	(0.106)	(0.105)	(0.121)	(0.119)
No line and $dist \in (73.4km, 151.5km]$	0.390***	0.388***	0.490***	0.500***
	(0.103)	(0.103)	(0.117)	(0.116)
No line and $dist \in (151.5km, 290.8km]$	0.154*	0.159*	0.162	0.180*
	(0.103)	(0.102)	(0.118)	(0.116)
Constant	3.264***	3.165***	2.827***	2.677***
	(0.076)	(0.080)	(0.084)	(0.088)
Observations	3,105	3,105	3,105	3,105
Controls		X		X

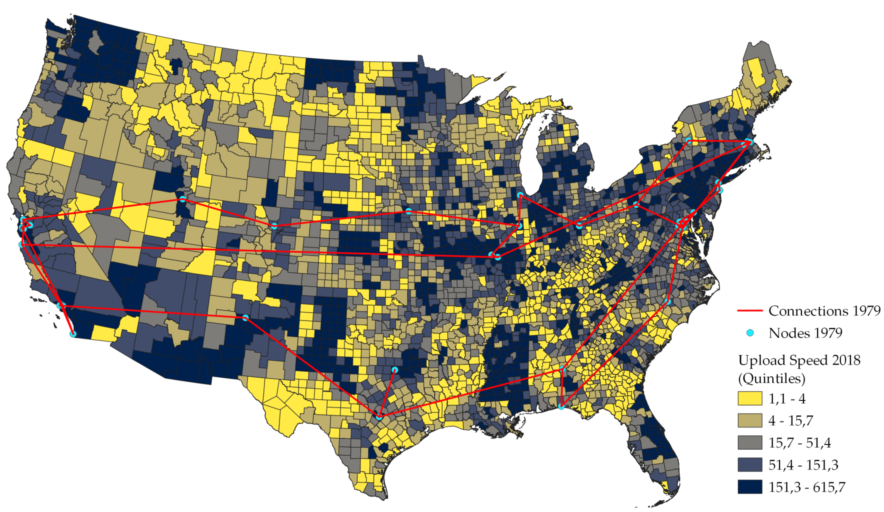
Notes: This table shows the impact of ARPANET in 1979 structure on today's internet, both the mean download and the mean upload speeds. The dependent variables are: a dummy variable that equals 1 if a county has a node (Panel A); a dummy variable that equals 1 if a county has a connection line (Panel B); the log of the distance between the country's centroid and a connection line (Panel C); and distance quartiles from a connection line (Panel D). In this last panel, counties between 290.8km and 1,118km belong to the omitted category. Controls include dummy variables for border and coastal counties, 1980 population and 1980 sectoral employment shares. SHAC adjusted standard errors (**?**) are in parenthesis. (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$)

Figure 3: ARPANET Connections in 1979 and 2018 Internet Speeds

Panel A: Download Speeds



Panel B: Upload Speeds



Note: this figure shows

Furthermore, the structure of ARPANET as an instrumental variable satisfies the exclusion restriction. In other words, ARPANET exclusively impacts the local economic structure in U.S. counties through modern Internet speeds. This is explained by the history of the foundation of the network. It is unlikely that the nodes chosen by the DoD through contractor processes, academic computing research agendas (interested in computer networks or not), and research work styles (collaborative or uncollaborative) are related to productivity or amenities in modern counties today. The agendas and how collaborative computer science departments were depends on the history of their academic departments. The contracting relationships depend on individual research networks. Moreover, national defense computing research locations were decided purely based on the interests of ARPANET. To give support to this explanation, we test that the ARPANET network in 1979 is unrelated to counties' economic structure in the 1970. Our results are YYYY

Table 3: test that ARPANET 1979 nodes does not influence 1970s sectoral shares (TBA).

Lastly, our instrumental variable is unlikely to be correlated with the omitted variables that influence the local economic structure, such as productivity or amenities; that is, the instrument satisfies the exogeneity assumption. This is supported by the fact that we use the ARPANET lines connecting the nodes, instead of the nodes themselves. Even if there are remaining concerns that the ARPANET nodes are located in counties with high amenities or productivity, the lines that connect them are likely to be exogenous to such unobserved factors. For a fictional example, an ARPANET line connects the Argonne National Laboratory in Argonne, IL to the Wright-Patterson Air Force Base in Green County, OH. Such line passes over several Indiana counties, including Wabash County. Even if under the strong assumption that Argonne, IL or Green County, OH nodes were selected due to high productivity, the line passing over Wabash County, IN is exogenous to the productivity of the Indiana county. The lines themselves predict the location of the ARPANET backbone, which itself predicts the geographical position of the equipment that forms the modern Internet backbone.

Our last concern for identification is whether our strategy violates the Stable Unit Treatment Value Assumption (SUTVA). In our case, SUTVA implies that potential outcomes for a given county respond only to its own Internet quality and are unrelated to the treatment status of other counties. Due to engineering related reasons, the local provision of Internet does not depend on another county's provision because ISPs are in charge of the last mile and can target infrastructure at granular levels.[17]. Such targeting capacity by ISPs dissipate concerns about violations of SUTVA.

[17] The provision of Internet has some similarities to electricity or water, cases in which the provider decides the quality of the service in the last mile, hence the quality of the service provision can be targeted (counties, neighborhoods, blocks, buildings, etc.). For example, quality can vary across tracts within the same city or even across neighborhoods within the same county. Some illustrative cases are Detroit (Wayne county) today or Chattanooga

Although distance to the Internet backbone can similarly affect the costs of providing high-speed Internet in neighboring counties (in the same way that distance to electricity generators drive the costs of electricity provision).

5 Results

In this section, we present the main results of the paper. In particular, we show the impact that better provision of Internet has on local economic activity. Moreover, we show its impacts on local structural transformation. Finally, we present the results regarding the two main proposed mechanisms behind such impact.

5.1 Internet quality and local economic activity

To study the effect of a better provision of internet on local economic aggregates, we start by estimating equation (1) using as dependent variables the county's GDP, its total annual payroll, total employment and average wages, defined as the ratio of the two previous variables. We estimate such equation by two-stage least squares (2SLS), where in a first stage we estimate equation (2) using the (log) distance between a country's centroid and an ARPANET connection line in 1979. In Table 3 we present the results of such estimation, using the (log) mean download speed in 2018 in Panel A and the (log) mean upload speed in 2018 in Panel B. In the table we include two types of standard errors: robust, and spatial heteroskedasticity and autocorrelation consistent standard errors (Conley, 1999), together with the first-stage Kleibergen-Paap for weak instruments (Kleibergen and Paap, 2006).

The general conclusion of these results is that better internet provision leads to better economic outcomes and to a more vibrant economic environment. In particular, a 1% improvement in internet speeds leads to an increase of 1%-1.3% in local GDP, as shown in column 3. As all these specifications include the ARPANET instrument, together with the vector of geographic and 1980 economic characteristics, we consider these results to be causal. Moreover, as the first stage KP statistic for weak instruments shows, the distance to an ARPANET line in 1979 is not a weak instrument for current internet speeds. We find similar results when using annual payroll (column 2) or employment (column 3) as alternative measures of economic activity: 1% higher internet speeds lead to an increase of 1.1%-1.5% in annual payroll and of 1.1%-1.4% in total employment.

The improvement in economic activity brought by better internet also makes workers better off, as

(Hamilton county) in the past, cities where national commercial ISPs did not provide high-speed Internet in specific neighborhoods, even though the rest of the city had access to broadband services. The importance of the last mile has been documented in detail by journalists, economists and non-profits (Lobo et al., 2008; Peñarroyo et al., 2022; Thornton and Mars, 2022)

Table 3: Effect of Better Internet Speed on Local Economic Aggregates

Panel A	Mean download speed			
	GDP	Annual Payroll	Employment	Average wage
	(1)	(2)	(3)	(4)
Log(Mean Download Speed)	1.313***	1.507***	1.424***	0.083***
	(0.182)	(0.209)	(0.196)	(0.021)
	[0.220]	[0.256]	[0.240]	[0.026]
Constant	8.976***	6.777***	3.438***	3.339***
	(0.631)	(0.730)	(0.685)	(0.073)
	[0.754]	[0.884]	[0.828]	[0.088]
FS F-Test	36.957	36.067	36.067	36.067
Panel B	Mean upload speed			
	GDP	Annual Payroll	Employment	Average wage
	(1)	(2)	(3)	(4)
Log(Mean Upload Speed)	1.003***	1.147***	1.084***	0.063***
	(0.126)	(0.144)	(0.134)	(0.016)
	[0.153]	[0.178]	[0.165]	[0.019]
Constant	10.465***	8.496***	5.063***	3.433***
	(0.382)	(0.442)	(0.414)	(0.048)
	[0.455]	[0.537]	[0.502]	[0.058]
FS F-Test	48.558	47.712	47.712	47.712
Observations	3,051	3,102	3,102	3,102
Controls	X	X	X	X

Notes: This table includes ... Average wages are measure as the ratio between total payroll and total employment. Robust standard errors are in parentheses and SHAC adjusted standard errors (**?**) are in brackets, with p-values corresponding to (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$)

seen by the positive impact on average wages shown in column 4. In particular, a 1% improvement in speeds leads to an increase in wages between 0.06% and 0.08%. To put this number into context, consider the percentiles 25th and 75th of the distribution of mean download speeds: 11.9 and 134.5, respectively, corresponding to a ten-fold difference (1000%). This improvement in internet would imply an increase of around $33,000 annually per worker. These results differ from those presented by Forman et al. (2012), who find a null impact of better internet on average wages. As discussed in the introduction, our results do not contradict theirs as they analyze the effects of internet in economic outcomes of the late 1990s, thus, focusing on the short-run effects of the initial internet boom.

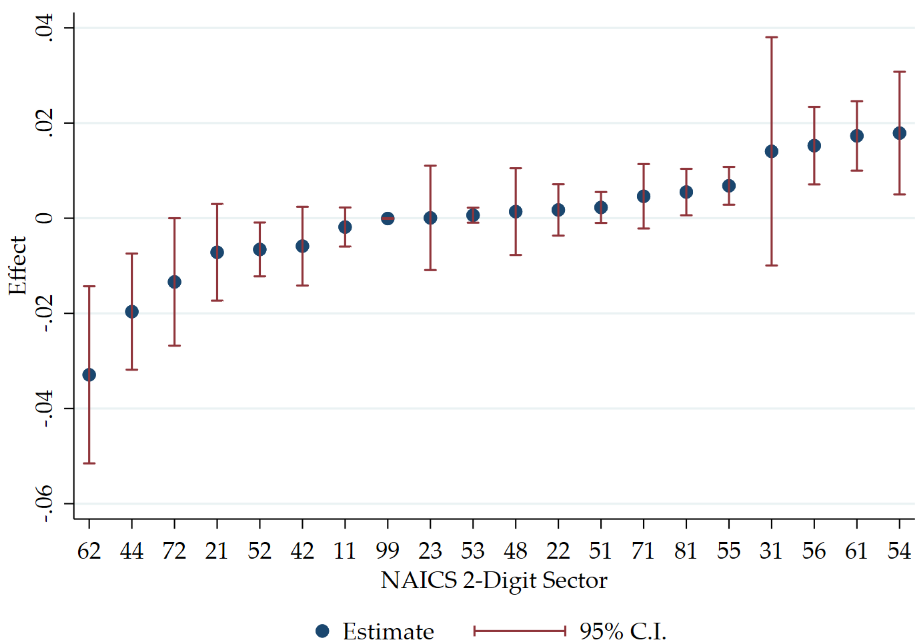
One reason explaining the large magnitude of these estimates can be that we are pooling together both urban and rural counties. We are currently working on re-estimating these equations including size-categories based on current population.

5.2 Internet quality and local structural transformation

We are also interested in understanding if and how a better provision of communications infrastructure alters the sectoral economic structure of countries, that is, if it leads to local structural transformation. Since research has already shown that better communication infrastructure can lead to a better transmission of ideas and innovation, it would be natural that it also favors sectors that can benefit the most from these innovations and from lower communication costs in general.

To explore local structural transformation, we estimate equation (1) using as dependent variable the share of employment in each sector using the (log) distance to an ARPANET connection line in 1979 as the IV. We start by analyzing these employment shares at the 2-digit NAICS sector. We present the results of these regressions in Figure 4, where we present the point estimate and their 95% confidence interval. The horizontal axis of this figure presents the 2-digit NAICS sector codes sorted by the size of the point estimate; the full name of each sector can be found in Table A-1.

Figure 4: Internet and Local Structural Transformation - Employment Shares



Note: this figure shows the causal effect of better internet provision (measured as higher download speeds) on the share of employment on different 2-digit NAICS sectors, estimated using the distance to an ARPANET line in 1979 as an instrument, together with geographic and 1980 economic characteristics. The horizontal axis shows the sectors sorted by the size of the estimate. The sector names are shown in Table A-1.

Results from Figure 4 show that better internet download speeds have a positive impact in the employment shares of four aggregate sectors: (i) Professional, Scientific, and Technical Services (54); (ii) Educational Services (61); (iii) Administrative and Support and Waste Management and Remediation Services (56); and (iv) Management of Companies and Enterprises (55). Manufacturing (31) also appears to have a positive point estimate, but it is quite imprecisely estimated, which could be explained by the large heterogeneity across manufacturing subsectors in this category. Notice that these positive effects are concentrated in what recent research has called Skilled-Scalable Services (Eckert et al., 2020) or Prime Services (Ahlfeldt et al., 2020), which are services intensive in high-skilled workers and information services, and have been shown to be responsible for the relative faster growth of larger cities and their structure.

On the other hand, faster download speeds negatively affects employment shares in four aggregate sectors: (i) Health Care and Social Assistance (62); (ii) Retail Trade (44); (iii) Accommodation and Food Services (72); and (iv) Finance and Insurance (52). These effects are statistically significant at the 95% confidence level. Notice that these four sectors can be considered to be present in basically every town and city across the country, which could suggest that regions with worse quality of internet might suffer from a lack of specialization.

The previous rankings are quite consistent if we use payroll shares instead of employment shares (Figure A-4) of if we use employment shares, together with the 1988 ARPANET structure as instrument (Figure A-5).

In also to capture what is going on inside each of these aggregate sectors, we re-estimated the model using employment shares for the 92 3-digit NAICS sectors. Given the large dimensionality of these results, we report them separately for Agriculture, Mining and Manufacturing in Table 4, and for all Services in Table 5. Moreover, we categorize all 3-digit sectors depending on whether their point estimate is significant at the 95% confidence level. The whole table containing the estimates and the standard errors for each subsector is available in the online appendix.

Regarding agriculture and manufacturing, notice two things. First, none of the manufacturing subsectors are negatively affected by better internet, and just five of them are positively impacted. In particular textile miles, plastics and rubber products, metal, electrical equipment, appliances and components are positively affected by better download speeds. Second, none of the agricultural or mining subsectors are positively affected by better internet.

Regarding services, notice three things. First within financial services,

TO BE COMPLETED

Table 4: Impacts of Higher Quality of Internet Provision on Agriculture and Manufacturing

Positive Impacts (95%)	**Zero Impacts**	**Negative Impacts (95%)**
313 - Textile Mills 326 - Plastics & Rubber Products Manufacturing 332 - Fabricated Metal Product Manufacturing 335 - Electrical Equip, Appliance, & Component Manuf. 339 - Miscellaneous Manuf.	113 Forestry & Logging 115 - Support Activities for Agriculture & Forestry 212 - Mining (except Oil & Gas) 213 - Support Activities for Mining 221 - Utilities 236 - Construction of Buildings 237 - Heavy & Civil Engineering Construction 238 - Specialty Trade Contractors 311 - Food Manufacturing 312 - Beverage & Tobacco Product Manufacturing 314 - Textile Product Mills 315 - Apparel Manufacturing 316 - Leather & Allied Product Manufacturing 322 - Paper Manufacturing 323 - Printing & Related Support Activities 324 - Petroleum & Coal Products Manufacturing 325 - Chemical Manufact. 327 - Nonmetallic Mineral Product Manufacturing 331 - Primary Metal Manufacturing 333 - Machinery Manufact. 334 - Computer & Electronic Product Manufacturing 336 - Transportation Equipment Manufacturing 337 - Furniture & Related Product Manufacturing	114 - Fishing, Hunting & Trapping 211 - Oil & Gas Extraction

5.3 Mechanism 1: Input-Output Linkages

Using the Input-Output matrix from the Bureau of Economic Analysis, we categorize each sector into quartiles based on its purchases from other ICT sectors, in particular, from *Broadcasting and Telecommunication* and *Information Services and Data Processing Services*.

Sectors with higher absolute or relative purchases from other ICT-intensive sectors are driving the results as shown in Table 6 where we use share of employment as the dependent variable. Table A-5 and Table A-6 show that these results are robust to the use to payroll, instead of employment, or tertiles or deciles, instead of quartiles.

TO BE COMPLETED

5.4 Mechanism 2: ICT driven Occupations

Local structural transformation could be driven by an inflow of workers in occupations that might benefit from the use of information and communication technologies. For instance, firms in sectors that benefit the most from higher Internet quality, such as business and administrative services, could decide to locate in place with a relatively higher abundance of workers in ICT driven occupations, e.g., computer scientists and financial support workers. This force can create a virtuous cycle in which these growing sectors end up attracting more ICT driven workers.

To test this hypothesis, we use data from the National Historical Geographic Information System (NHGIS) containing the number of workers in different occupation categories for each county. For each location, we compute the number and the share of workers in occupations that are more prone to using internet. In particular, we include the following occupational categories: management, business, and financial occupations, computer and mathematical occupations, architecture and engineering occupations, and office and administrative support occupations. Using these data, we estimate a regression equivalent to those used to obtain the results in Section 5. Specifically, we regress the log (and the share) of workers in ICT driven occupations on current Internet quality, instrumented using the distance of a county to an ARPANET connection line.

The results of this regression suggest that better Internet provision in a county increases the number of workers that work in occupations that might benefit from the use of ICT technologies. In particular, estimates from Panel A in Table 7 imply that a 1% increase in download speeds lead to a 1.3% increase in the number of workers in ICT-intensive occupations. These results also hold if, instead of occupations, we consider those workers with an educational attainment higher than a bachelor's degree: a master's, a professional school or a doctorate degree. Results from Panel B suggest that a 1% increase in download speeds lead to a 1.7% increase in the number of workers with more than a bachelor's degree; these results also show that the relative importance of these workers increase within the county. Table A-7 in the appendix show that these results are robust when we use upload speeds, instead of download speeds, to measure the quality of internet provision.

6 Conclusions

Higher quality communication infrastructure is important. Differently from other types of infrastructure, better CI can increase the transmission speed of ideas between individuals at relatively low costs. Internet in particular, can also improve the technical functionality of technological devices and enhance their commercial uses. Research has also shown that better CI can lead to more innovation and entrepreneurship, larger firms, higher housing prices and higher trade flows. However, their effects on local economies and local structural transformation had remained understudied.

In this paper, we document how differences in the quality of communication infrastructure influence the structural transformation of local economies. In particular, we use economic and internet data for all counties in the US in 2018, to explore the relationship between better internet, local economic outcomes and sectoral employment shares. For identification, we use the distance from each county to one of the lines connecting ARPANET nodes, a network that was the precursor of the internet and later became the backbone of Internet in its initial stage. These connection lines represent the actual telecommunications equipment installed to connect the network nodes. We obtain such distances from historical government reports documenting the early history of Internet. We combine these historical data with different geographic and 1980 economic characteristics.

Our estimates suggest that if a county improves its Internet quality, its GDP increases, as well as its total employment and average wages. Moreover, better internet favours employment in high-skilled services, such as, management, information, professional services and educational services. Nonetheless, better internet also leads to a decrease in other sectors, such as, retail, accommodation and food services, health care, and financial and insurance. Even though the negative effect on the financial sector might seem puzzling, they appear natural when we explore a more disaggregated sectoral structure. Specifically, better Internet reduces the county employment in *Credit Intermediation and Related Activities*, which mostly includes physical banks, while it increases employment in subsectors related with high-tech finance.

Our results have clear implications for infrastructure policy and public expenditures, as they suggest that Internet quality can explain regional inequality patterns. All types of countries have spent a large amount trying to improve the quality of communications infrastructure. Examples include advanced economies like Canada (Government of Canada, 2019), the United States (The White House, 2022), Germany (European Commission, 2022) and the UK (Hutton, 2022), and middle income economies like Colombia or Brazil. Even though not all these efforts have been successful, the motivation behind these policies is to reduce the technology education gap.

Table 5: Impacts of Higher Quality of Internet Provision on Services

Positive Impacts (95%)	**Zero Impacts**	**Negative Impacts (95%)**
511 - Publishing Industries (except Internet)	512 - Motion Picture & Sound Recording Industries	522 - Credit Intermediation & Related Activities
517 - Telecommunications	515 - Broadcasting (except Internet)	624 - Social Assistance
518 - Internet Service Providers, Web Search Portals, and Data Processing	521 - Monetary Authorities - Central Bank	424 - Merchant Wholesalers, Non-durable Goods
519 - Other Information	524 - Insurance Carriers & Related Activities	444 - Building Material, Garden Equipment & Supplies
523 - Securities, Commodity Contracts, & Other Financial Investments	525 - Funds, Trusts, & Other Financial Vehicles	447 - Gasoline Stations
551 - Management of Companies & Enterprises	531 - Real Estate	484 - Truck Transportation
611 - Educational Services	532 - Rental & Leasing Services	721 - Accommodation
425 - Wholesale Electronic Markets & Agents & Brokers	533 - Lessors of Non-financial Intangible Assets	722 - Food Services & Drinking Places
442 - Furniture & Home Furnishings Stores	541 - Professional, Scientific, & Technical Services	
443 - Electronics & Appliances Stores	621 - Ambulatory Health Care Services	
448 - Clothing & Clothing Accessories Stores	622 - Hospitals	
451 - Sporting Goods, Hobby, Book, & Music Stores	623 - Nursing & Residential Care Facilities	
485 - Transit & Ground Passenger Transportation	423 - Merchant Wholesalers, Durable Goods	
493 - Warehousing & Storage	441 - Motor Vehicle & Parts Dealers	
561 - Administrative & Support Services	445 - Food & Beverage Stores	
711 - Performing Arts, Spectator Sports, & Related	446 - Health & Personal Care Stores	
812 - Personal & Laundry Services	452 - General Merchandise Stores	
	453 - Miscellaneous Store Retailers	
	454 - Non-store Retailers	
	481 - Air Transportation	
	483 - Water Transportation	
	486 - Pipeline Transportation	
	487 - Scenic & Sightseeing Transportation	
	488 - Support Activities for Transportation	
	492 - Couriers & Messengers	
	562 - Waste Management & Remediation Services	
	712 - Museums, Historical Sites, & Similar Institutions	
	713 - Amusement, Gambling, & Recreation Industries	
	811 - Repair & Maintenance	
	813 - Religious, Civic, & Similar Organizations	

Table 6: Heterogeneity Analysis: Importance of Internet within Sector (by Quartiles) - Dependent Variable: Log Employment

Panel A	Absolute relevance - download speed			
	Quartile 4	Quartile 3	Quartile 2	Quartile 1
	(1)	(2)	(3)	(4)
Log(Mean download speed)	-0.001	0.014	-0.044**	0.051***
	(0.007)	(0.012)	(0.018)	(0.017)
Constant	0.066***	0.143***	0.766***	0.107*
	(0.024)	(0.042)	(0.066)	(0.060)
Panel B	Absolute relevance - upload speed			
	Quartile 4	Quartile 3	Quartile 2	Quartile 1
	(1)	(2)	(3)	(4)
Log(Mean upload speed)	-0.001	0.010	-0.034**	0.038***
	(0.005)	(0.009)	(0.014)	(0.013)
Constant	0.065***	0.159***	0.715***	0.166***
	(0.016)	(0.028)	(0.045)	(0.039)
Panel C	Relative relevance - download speed			
	Quartile 4	Quartile 3	Quartile 2	Quartile 1
	(1)	(2)	(3)	(4)
Log(Mean download speed)	0.006	-0.032*	-0.018	0.063***
	(0.007)	(0.016)	(0.015)	(0.021)
Constant	0.046*	0.371***	0.561***	0.105
	(0.026)	(0.060)	(0.052)	(0.074)
Panel D	Relative relevance - upload speed			
	Quartile 4	Quartile 3	Quartile 2	Quartile 1
	(1)	(2)	(3)	(4)
Log(Mean upload speed)	0.005	-0.024*	-0.014	0.048***
	(0.006)	(0.012)	(0.011)	(0.016)
Constant	0.053***	0.334***	0.539***	0.178***
	(0.018)	(0.040)	(0.036)	(0.048)
Observations	3,085	3,085	3,085	3,085
SE	Conley 50	Conley 50	Conley 50	Conley 50
Controls	X	X	X	X

Notes: This table shows ... SHAC adjusted standard errors (**?**) are in parenthesis. (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$)

Table 7: The Impact of Internet on the prevalence of ICT-intensive workers

Panel A	Workers in ICT related occupations	
	Share	Logs
	(1)	(2)
Log (Mean Download Speed)	0.012	1.342***
	(0.008)	(0.224)
Constant	0.241***	3.038***
	(0.026)	(0.774)
Panel B	Workers with more than a Bachelor's Degree	
	Share	Logs
	(1)	(2)
Log (Mean Download Speed)	0.039***	1.652***
	(0.009)	(0.273)
Constant	1.652***	0.941
	(0.273)	(0.947)
Observations	3,103	3,103
SE	Conley p50	Conley p50
Controls	X	X

Notes: This table shows ... SHAC adjusted standard errors (**?**) are in parenthesis. (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$)

References

Abbate, J. (2000). *Inventing the internet*. MIT press.

Acosta, C. and Lyngemark, D. H. (2020). Spatial wage differentials, geographic frictions, and the organization of labor within firms. *EAFIT University Working Paper*.

Ahlfeldt, G., Koutroumpis, P., and Valletti, T. (2017-07-01). Speed 2.0: Evaluating access to universal digital highways. *Journal of the European Economic Association*, 15(3):586–625.

Ahlfeldt, G. M., Albers, T. N., and Behrens, K. (2020). Prime locations. Working paper.

Allen, T. (2014). Information frictions in trade. *Econometrica*, 82(6):2041–2083.

Arfi, W. B. and Hikkerova, L. (2021). Corporate entrepreneurship, product innovation, and knowledge conversion: the role of digital platforms. *Small Business Economics*, 56(3):1191–1204.

Beem, R. (2022). Broadband internet and business activity. *University of Tennesse Working Paper*.

Beracha, E. and Wintoki, M. B. (2013). Forecasting residential real estate price changes from online search activity. *Journal of Real Estate Research*, 35(3):283–312.

Beyers, W. B. (2002). Services and the new economy: elements of a research agenda. *Journal of Economic Geography*, 2(1):1–29.

Blum, B. S. and Goldfarb, A. (2006). Does the internet defy the law of gravity? *Journal of International Economics*, 70(2):384–405.

Breinlich, H. and Criscuolo, C. (2011). International trade in services: A portrait of importers and exporters. *Journal of International Economics*, 84(2):188–206.

Bresnahan, T. F., Davis, J. P., and Yin, P.-L. (2014). Economic value creation in mobile applications. In *The changing frontier: Rethinking science and innovation policy*, pages 233–286. University of Chicago Press.

Carlino, G. A., Chatterjee, S., and Hunt, R. M. (2007). Urban density and the rate of invention. 61(3):389–419.

Cavallo, A. (2018). More amazon effects: online competition and pricing behaviors. Technical report, National Bureau of Economic Research.

Cerf, V. and Khan, R. (1990). Selected arpanet maps 1969-1990. *SIGCOMM Computer Communication Review*, 20(5).

Charlot, S. and Duranton, G. (2006). Cities and workplace communication: Some quantitative french evidence. *Urban Studies*, 43(8):1365–1394. _eprint: https://doi.org/10.1080/00420980600776459.

Conley, T. G. (1999). Gmm estimation with cross sectional dependence. *Journal of Econometrics*, 92(1):1–45.

Cristea, A. D. (2011). Buyer-seller relationships in international trade: Evidence from u.s. states' exports and business-class travel. *Journal of International Economics*, 84(2):207–220.

DARPA (1981). A History of the ARPANET: The First Decade. Technical Report 4799, Defense Advanced Research Projects Agency, 1400 Wilson Bldv, Arlington, VA, 22209.

DeStefano, T., Kneller, R., and Timmis, J. (2022-04). The (fuzzy) digital divide: the effect of universal broadband on firm performance*. *Journal of Economic Geography*. _eprint: https://academic.oup.com/joeg/advance-article-pdf/doi/10.1093/jeg/lbac006/43438904/lbac006.pdf.

Dietzel, M. A. (2016). Sentiment-based predictions of housing market turning points with google trends. *International Journal of Housing Markets and Analysis*.

Dingel, J. I. and Neiman, B. (2020). How many jobs can be done at home? *Journal of Public Economics*, 189:104235.

Duranton, G. (2015). Roads and trade in colombia. *Economics of Transportation*, 4(1):16–36.

Duranton, G., Morrow, P., and Turner, M. (2014). Roads and Trade: Evidence from the U.S. *Review of Economic Studies*, 81(2):681–724.

Eckert, F., Ganapati, S., and Walsh, C. (2020). Skilled scalable services: The new urban bias in economic growth. *Available at SSRN 3736487*.

Encyclopedia.com (2022). Node.

European Commission (2022). Broadband in germany.

Federal Communications Commission (2019). FCC form 477 local telephone competition and broadband reporting. instructions for filings as of december 31, 2019 and beyond.

Fink, C., Mattoo, A., and Neagu, I. C. (2005). Assessing the impact of communication costs on international trade. *Journal of International Economics*, 67(2):428–445.

Ford, J. S., Rutherford, R. C., and Yavas, A. (2005a). The effects of the internet on marketing residential real estate. *Journal of Housing Economics*, 14(2):92–108.

Ford, J. S., Rutherford, R. C., and Yavas, A. (2005b). The effects of the internet on marketing residential real estate. *Journal of Housing Economics*, 14(2):92–108.

Forman, C., Goldfarb, A., and Greenstein, S. (2012). The internet and local wages: A puzzle. *American Economic Review*, 102(1):556–75.

Forman, C. and Van Zeebroeck, N. (2012). From wires to partners: How the internet has fostered r&d collaborations within firms. *Management Sciences*, 58:1549–1568.

Freund, C. L. and Weinhold, D. (2004). The effect of the internet on international trade. *Journal of International Economics*, 62(1):171–189.

Gaspar, J. and Glaeser, E. L. (1998). Information technology and the future of cities. *Journal of Urban Economics*, 43(1):136–156.

Glaeser, E. L. and Ponzetto, G. A. (2007). Did the death of distance hurt detroit and help new york? Technical report, National Bureau of Economic Research.

Glaeser, E. L., Saiz, A., Burtless, G., and Strange, W. C. (2004). The rise of the skilled city. *Brookings-Wharton Papers on Urban Affairs*, pages 47–105. Publisher: Brookings Institution Press.

Government of Canada (2019). High-speed access for all: Canada's connectivity strategy.

Greenstein, S. and Spiller, P. (1996). Estimating the welfare effects of digital infrastructure.

Hardy, A. P. (1980). The role of the telephone in economic development. *Telecommunications Policy*, 4(4):278–286.

Harris, S. and Gerich, E. (1996). Retiring the NSFNET backbone service: Chronicling the end of an era. *ConneXions*, 10(4).

Hauben, M., Hauben, R., and Truscott, T. (1998). *Behind the Net: The Untold Story of the ARPANET and Computer Science.* Wiley-IEEE Computer Society Press.

Hutton, G. (2022). The universal service obligation (USO) for broadband.

Jiang, X. (2022). Information and communication technology and firm geographic expansion. *Duke University Working Paper.*

Juhász, R. and Steinwender, C. (2018). Spinning the web: Codifiability, information frictions and trade. *University of British Columbia Working Paper.*

Kantor, S. and Whalley, A. (2019). Research proximity and productivity: Long-term evidence from agriculture. *Journal of Political Economy*, 127(2):819–854. _eprint: https://doi.org/10.1086/701035.

Kleibergen, F. and Paap, R. (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics*, 133(1):97–126.

Leiner, B., Cerf, V., Clark, D., Kahn, R., Kleinrock, L., Lynch, D., Postel, J., Roberts, L., and Wolff, S. (1997). A brief history of the internet.

Licklider, J. C. and Taylor, R. W. (1968). The computer as a communication device. *Science and Technology*, 76(2):1–3.

Lin, J. (2011-05-01). Technological adaptation, cities, and new work. *The Review of Economics and Statistics*, 93(2):554–574.

Lobo, B., Novobilski, A., and Ghosh, S. (2008). The Economic Impact Of Broadband: Estimates From A Regional Input-Output Model. *Journal of Applied Business Research*, 24(2).

Malecki, E. J. (2002). The economic geography of the internet's infrastructure. *Economic Geography*, 78(4):399–424. Publisher: Routledge _eprint: https://www.tandfonline.com/doi/pdf/10.1111/j.1944-8287.2002.tb00193.x.

Marinoni, A. and Roche, M. (2022). You've got mail! communication infrastructure, firm entry and performance - evidence from the US postal service expansion 1880-1900. *Mimeo, Georgia Institute of Technology and Harvard University.*

McKenzie, A. and Walden, D. (1991). The arpanet, the defense data network, and the internet. *Froehlich, F. & Kent, A., Encyclopedia of Telecommunications. Marcel Dekker, New York*, pages 365–367.

Norton, S. W. (1992). Transaction costs, telecommunications, and the microeconomics of macroeconomic growth. *Economic Development and Cultural Change*, 41(1):175–196. Publisher: University of Chicago Press.

Oestmann, M. and Bennöhr, L. (2015). Determinants of house price dynamics. what can we learn from search engine data? *Review of Economics*, 66(1):99–127.

Peñarroyo, C., Lindquist, S., and Miller, R. (2022). Mapping detroit's digital divide. *University of Michigan Urban Laboratory*.

Rosenblat, T. S. and Mobius, M. M. (2004). Getting closer or drifting apart? *The Quarterly Journal of Economics*, 119(3):971–1009. Publisher: Oxford University Press.

Röller, L.-H. and Waverman, L. (2001). Telecommunications infrastructure and economic development: A simultaneous approach. *The American Economic Review*, 91(4):909–923. Publisher: American Economic Association.

Steinwender, C. (2018). Real effects of information frictions: "when the states and the kingdom became united". *American Economic Review*, 108(3):657–696.

The White House (2022). Biden-Hharris Administration Announces Over $25 Billion in American Rescue Plan Funding to Help Ensure Every American Has Access to High Speed, Affordable Internet.

Thornton, K. and Mars, R. (2022). The future of the final mile. *Podcast 99 Percent Invisible, Episode 481*.

U.S. Congress (1992). Management of NSFNET. hearing before the subcommittee on science, space, and technology, u.s. house of representatives, 102nd congress, second session. page 190.

Wall, C. (2021). Invisible and vital: Undersea cables and transatlantic security. *Center for Strategic and International Studies*.

Zook, M. A. (2002). Grounded capital: venture financing and the geography of the internet industry, 1994–2000. 2(2):151–177.

Quality of Communications Infrastructure Provision and Local Structural Transformation

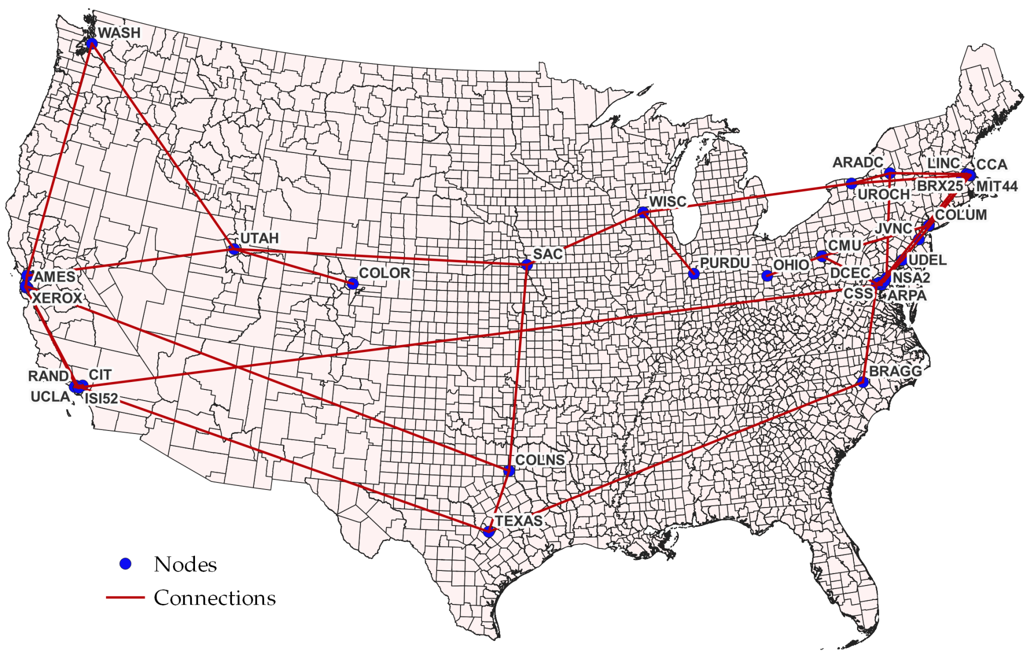
Camilo Acosta, Universidad EAFIT

Luis Baldomero-Quintana, William & Mary

Appendix
(for online publication)

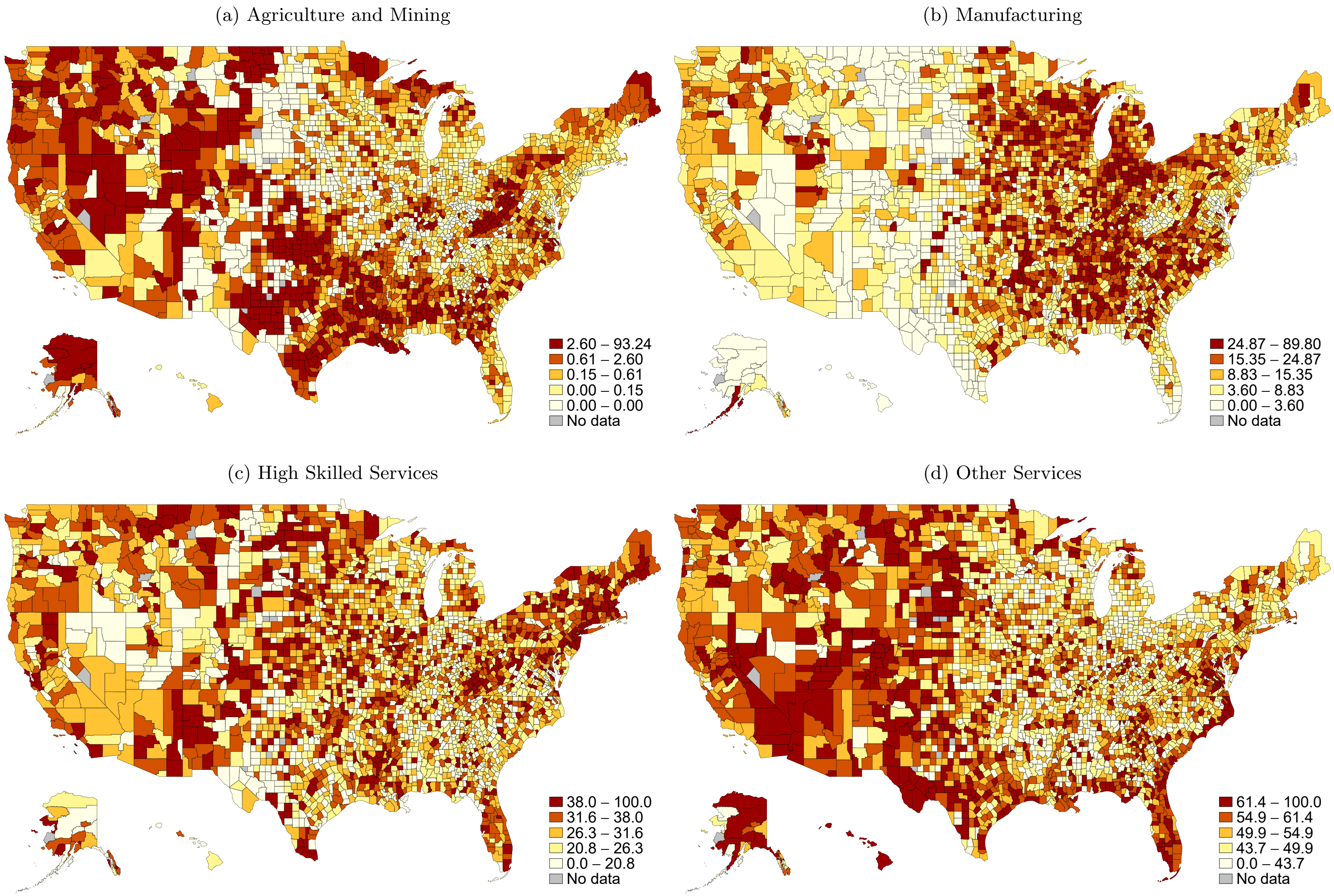
A1 Extra Figures and Tables

Figure A-1: ARPANET - 1988, Digitized Map



Note: this figure shows

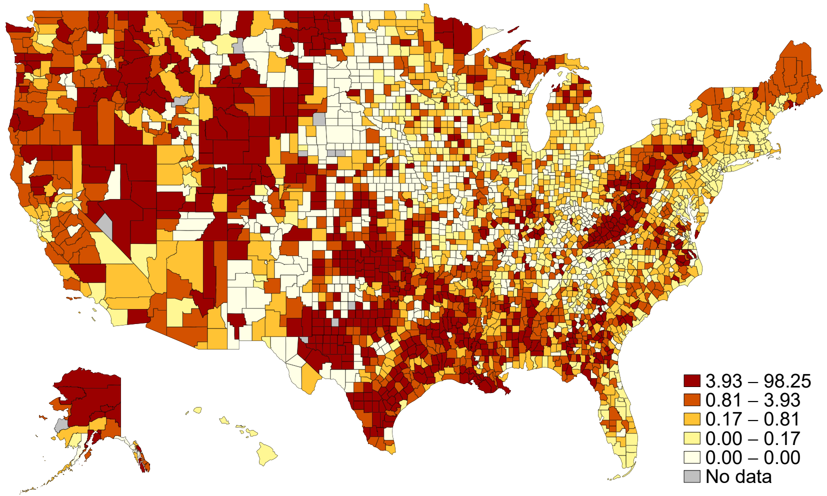
Figure A-2: Employment Shares by Aggregate Sectors



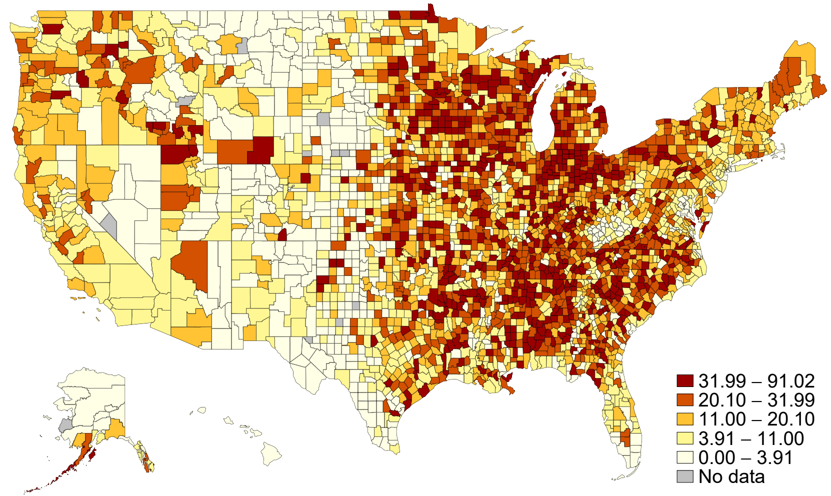
Note: these maps show the shares of employment in each county by quintiles using data from the 2018 County Business Patterns.

Figure A-3: Payroll Shares by Aggregate Sectors

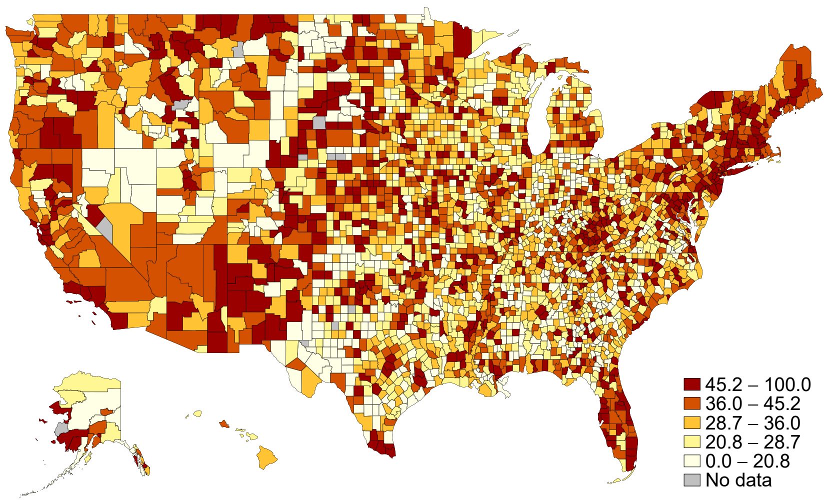
(a) Agriculture and Mining



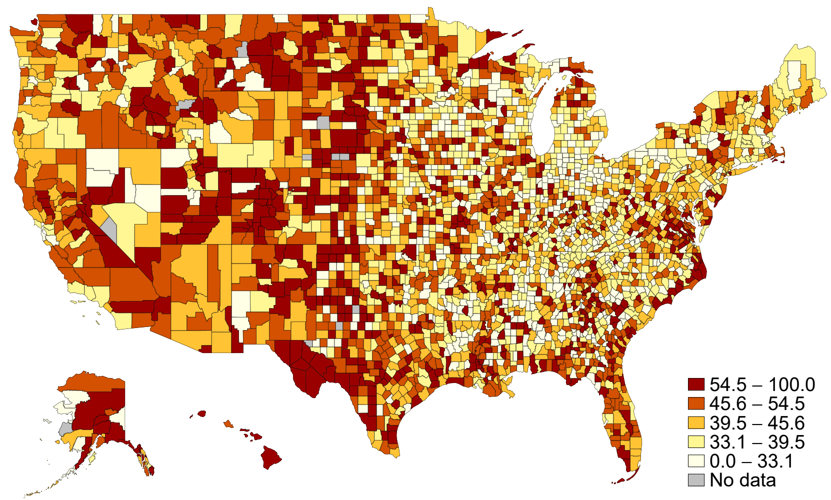
(b) Manufacturing



(c) High Skilled Services

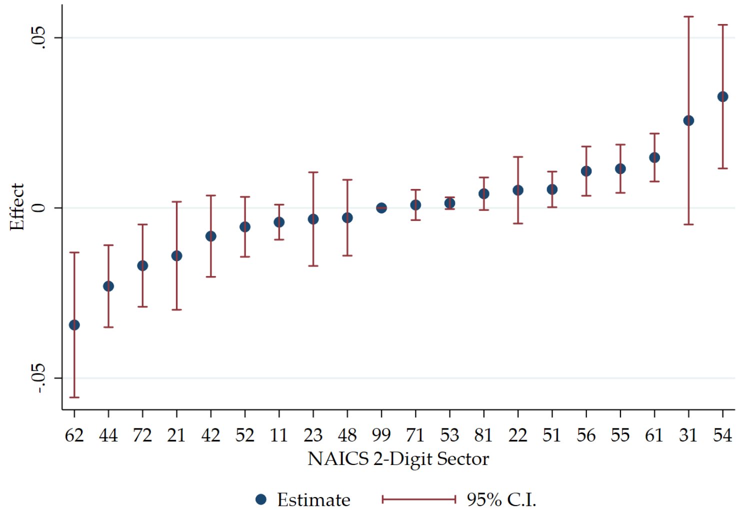


(d) Other Services



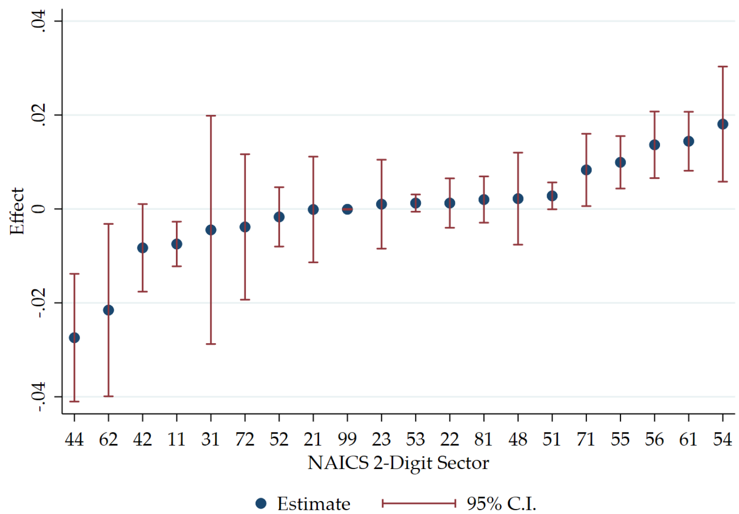
Note: these maps show the shares of aggregate payroll in each county by quintiles using data from the 2018 County Business Patterns.

Figure A-4: Internet and Local Structural Transformation - Payroll Shares



Note: this figure shows the causal effect of better internet provision (measured as higher download speeds) on the share of payroll on different 2-digit NAICS sectors, estimated using the distance to an ARPANET line in 1979 as an instrument, together with geographic and 1980 economic characteristics. The horizontal axis shows the sectors sorted by the size of the estimate. The sector names are shown in Table A-1.

Figure A-5: Internet and Local Structural Transformation - Employment Shares an IV ARPANET in 1988



Note: this figure shows the causal effect of better internet provision (measured as higher download speeds) on the share of employment on different 2-digit NAICS sectors, estimated using the distance to an ARPANET line in 1988 as an instrument, together with geographic and 1980 economic characteristics. The horizontal axis shows the sectors sorted by the size of the estimate. The sector names are shown in Table A-1.

Table A-1: List of NAICS 2-Digit Sectors

Sector	Description
11	Agriculture, Forestry, Fishing and Hunting
21	Mining, Quarrying, and Oil and Gas Extraction
22	Utilities
23	Construction
31-33	Manufacturing
42	Wholesale Trade
44-45	Retail Trade
48-49	Transportation and Warehousing
51	Information
52	Finance and Insurance
53	Real Estate and Rental and Leasing
54	Professional, Scientific, and Technical Services
55	Management of Companies and Enterprises
56	Administrative and Support and Waste Management and Remediation Services
61	Educational Services
62	Health Care and Social Assistance
71	Arts, Entertainment, and Recreation
72	Accommodation and Food Services
81	Other Services (except Public Administration)

Table A-2: The Impact of ARPANET in 1988 on Current Internet

***Panel A**: County has a node*	Mean download speed		Mean upload speed	
	(1)	(2)	(3)	(4)
Had node in 1988	1.433***	0.550**	1.738***	0.698***
	(0.161)	(0.227)	(0.171)	(0.258)
Constant	3.601***	3.504***	3.239***	3.105***
	(0.039)	(0.049)	(0.046)	(0.059)
***Panel B**: County has a line*	Mean download speed		Mean upload speed	
	(1)	(2)	(3)	(4)
Had connection line in 1988	0.334***	0.192**	0.438***	0.274***
	(0.080)	(0.081)	(0.092)	(0.093)
Constant	3.561***	3.481***	3.185***	3.071***
	(0.042)	(0.050)	(0.049)	(0.060)
***Panel C**: Distance to a line*	Mean download speed		Mean upload speed	
	(1)	(2)	(3)	(4)
Distance to Line in 1988	-0.175***	-0.142***	-0.235***	-0.201***
	(0.025)	(0.026)	(0.029)	(0.030)
Constant	5.600***	5.115***	5.923***	5.385***
	(0.285)	(0.299)	(0.329)	(0.345)
***Panel D**: Distance to a line*	Mean download speed		Mean upload speed	
(Categories)	(1)	(2)	(3)	(4)
Has a line	0.809***	0.676***	1.073***	0.935***
	(0.106)	(0.109)	(0.121)	(0.124)
No line and $dist \in (0km, 72.2km]$	0.710***	0.681***	0.976***	0.962***
	(0.097)	(0.097)	(0.110)	(0.110)
No line and $dist \in (72.2km, 131.3km]$	0.658***	0.661***	0.878***	0.899***
	(0.097)	(0.096)	(0.111)	(0.110)
No line and $dist \in (131.3km, 224.7km]$	0.530***	0.545***	0.685***	0.715***
	(0.099)	(0.097)	(0.114)	(0.112)
Constant	3.086***	2.999***	2.550***	2.412***
	(0.075)	(0.078)	(0.086)	(0.089)
Observations	3,105	3,105	3,105	3,105
Controls		X		X

Notes: This table shows the impact of ARPANET in 1988 structure on today's internet, both the mean download and the mean upload speeds. The dependent variables are: a dummy variable that equals 1 if a county has a node (Panel A); a dummy variable that equals 1 if a county has a connection line (Panel B); the log of the distance between the country's centroid and a connection line (Panel C); and distance quartiles from a connection line (Panel D). In this last panel, counties between 224.7km and 1,012.9km belong to the omitted category. Controls include dummy variables for border and coastal counties, 1980 population and 1980 sectoral employment shares. SHAC adjusted standard errors (?) are in parenthesis. (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$)

Table A-3: Effect of Better Internet Speed on Local Economic Aggregates
Other Standard Errors

Panel A	SHAC p75 Standard Errors			
	GDP (1)	Annual Payroll (2)	Employment (3)	Average wage (4)
Log(Mean Download Speed)	1.313***	1.507***	1.424***	0.083***
	(-0.182)	(-0.209)	(-0.196)	(-0.021)
	[-0.238]	[-0.278]	[-0.26]	[-0.027]
Constant	8.976***	6.777***	3.438***	3.339***
	(-0.631)	(-0.73)	(-0.685)	(-0.073)
	[-0.809]	[-0.954]	[-0.896]	[-0.094]
FS F-Test	31.20	30.83	30.83	30.83
Panel B	SHAC p90 Standard Errors			
	GDP (1)	Annual Payroll (2)	Employment (3)	Average wage (4)
Log(Mean Download Speed)	1.313***	1.507***	1.424***	0.083***
	(-0.182)	(-0.209)	(-0.196)	(-0.021)
	[-0.261]	[-0.307]	[-0.288]	[-0.03]
Constant	8.976***	6.777***	3.438***	3.339***
	(-0.631)	(-0.73)	(-0.685)	(-0.073)
	[-0.884]	[-1.053]	[-0.989]	[-0.103]
FS F-Test	25.74	25.42	25.42	25.42
Observations	3,051	3,102	3,102	3,102
Controls	X	X	X	X

Notes: This table includes ... Average wages are measure as the ratio between total payroll and total employment. Robust standard errors are in parentheses and SHAC adjusted standard errors (?) are in brackets, with p-values corresponding to (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$)

Table A-4: Effect of Better Internet Speed on Local Economic Aggregates, using 1988 ARPANET Structure

Panel A	Mean download speed			
	GDP	Annual Payroll	Employment	Average wage
	(1)	(2)	(3)	(4)
Log(Mean Download Speed)	1.704***	1.844***	1.695***	0.150***
	(0.239)	(0.264)	(0.242)	(0.029)
	[0.272]	[0.303]	[0.278]	[0.033]
Constant	7.613***	5.594***	2.489***	3.105***
	(0.837)	(0.928)	(0.850)	(0.102)
	[0.943]	[1.057]	[0.972]	[0.114]
F test	31.463	30.243	30.243	30.243
Panel B	Mean upload speed			
	GDP	Annual Payroll	Employment	Average wage
	(1)	(2)	(3)	(4)
Log(Mean Upload Speed)	1.204***	1.303***	1.198***	0.106***
	(0.143)	(0.161)	(0.147)	(0.019)
	[0.164]	[0.186]	[0.171]	[0.022]
Constant	9.845***	8.011***	4.710***	3.301***
	(0.439)	(0.495)	(0.454)	(0.059)
	[0.499]	[0.571]	[0.524]	[0.066]
F test	47.163	45.242	45.242	45.242
Observations	3,051	3,102	3,102	3,102
Controls	X	X	X	X

Notes: This table includes ... Average wages are measure as the ratio between total payroll and total employment. Robust standard errors are in parentheses and SHAC adjusted standard errors (**?**) are in brackets, with p-values corresponding to (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$)

Table A-5: Heterogeneity Analysis: Importance of Internet within Sector (by Quartiles) - Dependent Variable: Log Annual Payroll

Panel A	Absolute relevance - download speed			
	Quartile 4	Quatrile 3	Quartile 2	Quartile 1
	(1)	(2)	(3)	(4)
Log(Mean download speed)	-0.003	0.034*	-0.090**	0.058**
	(0.009)	(0.018)	(0.036)	(0.023)
Constant	0.093***	0.150**	0.858***	0.146*
	(0.030)	(0.062)	(0.148)	(0.084)
Panel B	Absolute relevance - upload speed			
	Quartile 4	Quatrile 3	Quartile 2	Quartile 1
	(1)	(2)	(3)	(4)
Log(Mean upload speed)	-0.002	0.026**	-0.068**	0.044**
	(0.007)	(0.013)	(0.027)	(0.018)
Constant	0.090***	0.189***	0.754***	0.213***
	(0.021)	(0.041)	(0.106)	(0.056)
Panel C	Relative relevance - download speed			
	Quartile 4	Quatrile 3	Quartile 2	Quartile 1
	(1)	(2)	(3)	(4)
Log(Mean download speed)	0.428	-0.081**	-0.014	0.084***
	(0.398)	(0.035)	(0.018)	(0.030)
Constant	2.037	0.650***	0.433***	0.106
	(1.404)	(0.144)	(0.066)	(0.107)
Panel D	Relative relevance - upload speed			
	Quartile 4	Quatrile 3	Quartile 2	Quartile 1
	(1)	(2)	(3)	(4)
Log(Mean upload speed)	0.324	-0.061**	-0.010	0.064***
	(0.301)	(0.026)	(0.014)	(0.022)
Constant	2.532***	0.557***	0.417***	0.203***
	(0.947)	(0.102)	(0.045)	(0.071)
Observations	3,085	3,085	3,085	3,085
SE	Conley 50	Conley 50	Conley 50	Conley 50
Controls	X	X	X	X

Notes: This table shows ... SHAC adjusted standard errors (?) are in parenthesis. (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$)

Table A-6: Heterogeneity Analysis: Importance of Internet within Sector (other classifications) - Dependent Variable: Log Employment

Panel A	Absolute relevance - download speed (employment)					
	Tertile 3	Tertile 2	Tertile 1	Decile 10	Decile 5	Decile 1
	(1)	(2)	(3)	(4)	(5)	(4)
Log(Mean download speed)	-0.001	0	0.032	0.007	0.013	0.072***
	(0.008)	(0.017)	(0.023)	(0.005)	(0.009)	(0.014)
Constant	0.133***	0.454***	0.495***	0.001	0.046	-0.154***
	(0.028)	(0.060)	(0.080)	(0.016)	(0.031)	(0.049)
Panel B	Relative relevance - download speed (employment)					
	Tertile 3	Tertile 2	Tertile 1	Decile 10	Decile 5	Decile 1
	(1)	(2)	(3)	(4)	(5)	(4)
Log(Mean download speed)	-0.009	-0.01	0.038	0.002	-0.028*	0.072***
	(0.009)	(0.017)	(0.023)	(0.004)	(0.016)	(0.014)
Constant	0.128***	0.466***	0.489***	0.012	0.304***	-0.154***
	(0.032)	(0.060)	(0.081)	(0.014)	(0.058)	(0.049)
Panel C	Absolute relevance - download speed (annual payroll)					
	Tertile 3	Tertile 2	Tertile 1	Decile 10	Decile 5	Decile 1
	(1)	(2)	(3)	(4)	(5)	(4)
Log(Mean download speed)	-0.003	-0.031	0.059*	0.01	0.030**	0.074***
	(0.011)	(0.035)	-0.031	-0.006	-0.013	-0.015
Constant	0.182***	0.667***	0.359***	0.003	0.016	-0.156***
	-0.038	-0.142	-0.111	-0.022	-0.045	-0.053
Panel D	Relative relevance - download speed (annual payroll)					
	Tertile 3	Tertile 2	Tertile 1	Decile 10	Decile 5	Decile 1
	(1)	(2)	(3)	(4)	(5)	(4)
Log(Mean download speed)	-0.01	-0.049	0.059*	0.003	-0.077**	0.074***
	(0.012)	(0.036)	(0.031)	(0.005)	(0.035)	(0.015)
Constant	0.167***	0.714***	0.366***	0.018	0.566***	-0.157***
	(0.042)	(0.144)	(0.111)	(0.019)	(0.142)	(0.053)
Observations	3,085	3,085	3,085	3,085	3,085	3,085
SE	Conley 50	Conley 50	Conley 50	Conley 50	Conley 50	Conley 50
Controls	X	X	X	X	X	X

Notes: This table shows ... SHAC adjusted standard errors (**?**) are in parenthesis. (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$)

Table A-7: The Impact of Internet on the prevalence of ICT-intensive workers, Upload Speeds

Panel A	Workers in ICT related occupations	
	Share	Logs
	(1)	(2)
Log (Mean Upload Speed)	0.009	1.022***
	(0.006)	(0.154)
Constant	0.254***	4.567***
	(0.018)	(0.467)
Panel B	Workers with more than a Bachelor's Degree	
	Share	Logs
	(1)	(2)
Log (Mean Upload Speed)	0.030***	1.258***
	(0.006)	(0.186)
Constant	-0.019	2.826***
	(0.020)	(0.568)
Observations	3,103	3,103
SE	Conley p50	Conley p50
Controls	X	X

Notes: This table shows ... SHAC adjusted standard errors (**?**) are in parenthesis. (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$)