When All Products Are Digital: Complexity and Intangible Value in the Ecosystem of Digitizing Firms

Pouya Rahmati, Ali Tafti, J Christopher Westland, César A. Hidalgo

When All Products Are Digital: Complexity and Intangible Value in the Ecosystem of Digitizing Firms

FORTHCOMING AT *MIS Quarterly*

Pouya Rahmati
Terry College of Business
University of Georgia
C425 Benson Hall, 630 S. Lumpkin St.
Athens, GA 30602
rahmati@uga.edu

Ali Tafti
College of Business Administration
University of Illinois at Chicago
601 S. Morgan St., Chicago, IL 60607
atafti@uic.edu

J. Christopher Westland
College of Business Administration
University of Illinois at Chicago
601 S. Morgan St., Chicago, IL 60607
westland@uic.edu

Cesar Hidalgo
ANITI, University of Toulouse
41 Allée Jules Guesde, 31000 Toulouse, France
cesar_hidalgo@univ-toulouse.fr

Abstract

During the last four decades, digital technologies have disrupted many industries. Car control systems have gone from mechanical to digital. Telephones have changed from sound boxes to portable computers. But have the firms that digitized their products and services become more valuable than firms that didn't? Here we introduce the construct of digital proximity, which considers the interdependent activities of firms linked in an economic network. We then explore how the digitization of products and services affects a company's Tobin's q—the ratio of market value over assets—a measure of the intangible value of a firm. Our panel regression methods and robustness tests suggest the positive influence of a firm's digital proximity on its Tobin's q. This implies that firms able to come closer to the digital sector have increased their intangible value compared to those that have failed to do so. These findings contribute a new way of measuring digitization and its impact on firm performance that is complementary to traditional measures of information technology (IT) intensity.

Keywords: complexity, intangible value, Tobin's q, digitization, software industry, pervasive computing, social networks

When All Products Become Digital: Complexity and Intangible Value in the Ecosystem of Digitizing Firms

INTRODUCTION

Fifty years ago, our cars were mechanical, our refrigerators electrical, and our buildings were not smart. Now, computing can be found almost everywhere, from the complex control systems of our cars' engines to the penny-sized computers driving our electric toothbrushes. The local mechanic can no longer repair the latest Toyota models without sophisticated and specialized software, while the 1990 models could be repaired in a home garage. This embedding of electronics and communication capabilities in everyday objects, called ubiquitous or pervasive computing, has evolved over four decades from a technical concept to an everyday reality.[1,2] Yet, pervasive computing is not just a technical curiosity; it is an economic phenomenon with implications for the performance and sustainability of firms.

The question we explore here is how does the integration of digital capacities into a firm's products and services affect its value? We explore this idea by examining the value of relatedness to digital industries.[3] We define relatedness as the distance between a firm and the digital sector and measure it by constructing a network connecting pairs of industries that firms are likely to participate in. We call this measure of relatedness *digital proximity*. To quantify digital proximity, we construct a network of industries. We use the sales that firms report in different industries to create a network representation of the industry space from 1990 to 2017.

---

[1] Satyanarayanan, M. 2001. "Pervasive Computing: Vision and Challenges," *IEEE Personal Communications* (8:4), pp. 10-17 (doi: 10.1109/98.943998).
[2] Weiser, M. 1993. "Some Computer Science Issues in Ubiquitous Computing," *Communications of the ACM* (36:7), pp. 75-84. Why are these references included here rather than as an in-text citation? I reformatted the reference here to match the style guide from MIS Q.
[3] Related measures are often used in the literature on economic complexity and economic geography to gauge the similarity of products (Hidalgo et al. 2007), industries (Neffke et al. 2011), research activities (Guevara et al. 2016), and patents (Boschma et al. 2015), as well as the countries, regions, and firms involved in such economic activities.

We connect two industries when many firms report sales in both of them, suggesting a relationship between the assets and capabilities required for operating in them. We then use this network to measure firms' proximity to the digital sector and explore whether firms with higher proximities are better valued in the market.

Digital proximity allows us to focus on the digitization of products and services, rather than the use of information technology (IT) for productivity or strategic agility. In contrast to traditional constructs (see Table 1 for a summary of the relevant literature), digital proximity has the following characteristics: (1) it is outcome-based, capturing the combined output of a firm's IT spending and capabilities, digital innovations, technology partnerships, and other contributing factors and mechanisms, and (2) it is network-based, capturing a firm's position toward digitization within a network of products and services. In this way, digital proximity helps explore digitization explicitly and complements other existing constructs.

For example, the leap in intensity of digitization for automobiles over the past four decades is an outcome brought about by the totality manufacturers' efforts toward digitization. Within a network of interdependent firms, the digitization of specific car components has shifted strategic resources from automobile manufacturers (e.g. Tesla, Toyota) and transportation services (e.g. Lyft) to producers of software and other digital technology (e.g. Google). Here, we point out that the traditional measures of IT intensity may not adequately capture the digital intensity of a firm's products. In the future, traditional IT intensity measures may become increasingly inadequate as digital platforms on cloud-computing ecosystems reduce the importance of monetary IT investment and increase the importance of the strategic positioning of digital capabilities. We use digital proximity to explore whether firms

become more valuable as they approach the digital sector. We focus on firms whose primary products are not digital; hence, we are interested in the digital intensity of products developed by *non-digital* firms. For a definition of digital (i.e. IT-producing) industries, see Table A1 of Appendix A.

We provide evidence that greater digital proximity leads to higher intangible value (higher Tobin's q) using various models that control for firm fixed-effects and other confounding factors to mitigate the potential for endogeneity. Our results contribute to the information systems (IS) literature in several ways. First, our results answer the recent call for research to explore IT beyond its impact on productivity and establish a link between firms' actual digitization levels and their performance (Bharadwaj et al. 2013). Second, our novel construct provides a new theoretical lens for exploring digitization as a distributed and generative phenomenon which expands beyond firm boundaries. Third, our proposed measure—while complementary to traditional measures of IT intensity—distinguishes firms that are more successful than others in converting knowledge and know-how into digital products. Accordingly, we will point out examples for which traditional measures of IT intensity fail to distinguish such firms while our proposed measure succeeds. Finally, digital proximity considers macro-economic trends of digitization in assessing each focal firm's success, thereby overcoming the data limitation issues that hinder research on the economic impacts of IT.

LITERATURE REVIEW

Digital Proximity and Economic Complexity

Our proposed construct of digital proximity is rooted in the literature on economic complexity. In a stream of recent studies, economic complexity represents the capacity of a nation's productive structure to develop diverse and distinctive

4

products (Hartmann et al. 2017; Hausmann et al. 2007; Hidalgo and Hausmann 2009; Hidalgo et al. 2007; Rodrik 2006). That is, to develop diverse and unique products, a country requires a productively functioning network of supporting institutions and norms, logistics and technological infrastructure, and coordinating processes to enable smooth market operations (Hartmann et al. 2017; Hausmann et al. 2007; Hidalgo 2015).

Products and services often require vastly more know-how than can be embodied in a single individual. As products become more sophisticated, they require more extensive social and professional networks to support their development (Chaney 2014; Hartmann et al. 2017; Hausmann and Hidalgo 2011; Hidalgo et al. 2007). A sophisticated economy coordinates the specialized know-how of vast numbers of skilled individuals within firms; in turn, firms are organized as coordinated networks of organizations. In much the same way that the human brain is a network that comprises an information-processing machine, a nation's economy is an agglomeration of networks forming a machine that processes information of enormous complexity to create sophisticated products and services (Hidalgo 2015).

Digital Proximity and Digitization

Prior research has studied the firm-value effects of IT (Melville et al. 2004). This literature uses a-priori or input-based notions of IT investment (Aral and Weill 2007; Brynjolfsson and Hitt 1996; Mithas et al. 2012; Rai et al. 1997), or the ability of firms to combine individual IT resources, such as IT infrastructure and human IT resources, into superior IT systems (Chae et al. 2014). In addition, this literature has viewed IT as predominantly back-office technologies that enhance firms' productivity in achieving strategic goals (Melville et al. 2004; Powell and Dent-Micallef 1997; Wade and Hulland 2004). While this approach has yielded important and significant

results, the commentary by Bharadwaj et al. (2013) calls for new measures that capture the growing complexity of digital innovations. Contemporary phenomena, such as pervasive computing, have transformed IT from back-office technologies to a component embedded in many products and services (Henfridsson and Bygstad 2013; Yoo et al. 2010). More recent IS literature acknowledges the reality of digital transformation and in turn explores the effect of digital innovations on firms' performance (Branstetter et al. 2019; Nan and Tanriverdi 2017; Nandkumar et al. 2018). Table 1 provides a summary of the relevant literature.

IS research posits that the digitization of products and services makes them more complex (Novales et al. 2016); specifically, digital products are modular (Yoo et al. 2010), enable product or service ecosystems (Barrett et al. 2015; Gawer 2014; Yoo et al. 2012), and result from generative and distributed innovations (Baldwin and Von Hippel 2011; Maznevski and Chudoba 2000; Parmar et al. 2014). Further, digitizing physical products introduces new challenges, including the management of generated data (Parmar et al. 2014), the complexity of supporting business processes (Mocker et al. 2014), and the economic ecosystem in which the products operate (Porter and Heppelmann 2015). Therefore, digitization enhances the sophistication of products and services—and intensifies the complexity and dynamics of the underlying social structures required for their development.

Digital proximity is an outcome-based socio-technological construct insofar as it is based on the specific position of firms in a network of related industries.[4] Firms develop digital products by coordinating multiple networks of assets, capabilities, systems, and stakeholders. A digitizing firm requires complex and evolving social

---

[4] Our definition of digital proximity is also in-line with the strategy literature showing the effect of geographic proximity on firms' innovation (Forman et al. 2016; Klepper 2010). Klepper (2010), in particular, shows that the geographic proximities developed in economic hubs are closely related to the business relatedness between firms in these areas, creating a leap between positioning in geographic networks and complementarity networks.

structures—managerial structures, business processes, and groups of professionals—to accumulate the know-how required for addressing the challenges rooted in the aforementioned complexities of digitization. While economic complexity entails a country's ability to produce sophisticated products, we posit digital proximity as the measure of a firm's capacity to: (i) develop digital products and services, (ii) embed digital components into products and services, and (iii) reconfigure and recombine know-how into new classes of products and services. Following prior studies on economic complexity, we develop the digital proximity construct to represent rent-generating digital capabilities embodied in industries that are not traditionally understood to be software-producing. We then hypothesize that firms' ability to develop such digital capabilities predicts higher levels of intangible value (measured using Tobin's q).

THEORETICAL FRAMEWORK

Resource-based theory suggests social complexity, path dependence, and environmental complexity as mechanisms that hinder competitors' ability to imitate firms' proximity to digital resources within a network of related industries (Barney 1991; Teece et al. 1997; Wade and Hulland 2004). Firms accumulate knowledge and know-how from their resources and capabilities (Amit and Schoemaker 1993; Dyer and Singh 1998; McGrath et al. 1996; Ray et al. 2004). This collective knowledge and know-how is specific to each firm, imperfectly mobile, and not easily imitable (Barney 1991; Bhatt and Grover 2005). Within a firm, collective knowledge and know-how often accumulate in networks of specialists and organizations; therefore, obtaining high levels of digital proximity entails high levels of social complexity. In addition, complex digital resources and capabilities are hard to imitate as firms' current level of proximity to digital industries in the industry space is the result of

7

their specific paths taken in the past (Eisenhardt and Martin 2000; Teece et al. 1997).

Firms that grow in digital proximity have created a combination of assets and

capabilities that serve as complements to digital assets and capabilities. Firms can

only develop such a combination of capabilities through a transformation process, in

the sense that these capabilities cannot simply be purchased (Brynjolfsson and Hitt

2000; Melville et al. 2004); rather, they are co-developed over time (Kohli and Grover

2008). Firms learn through their many attempts to identify and leverage new business

opportunities, and their learning processes become embodied in their unique set of

resources and capabilities (Teece et al. 1997). Finally, due to the fundamentally

unpredictable dynamics of the environment, digital proximity entails a complex

competitive environment in which a firm operates (Lee et al. 2010). In recent decades,

firms in many industries have experienced unexpected competition from IT or

software firms (Hacklin et al. 2013). Unwittingly, they have found themselves

operating in complex environments, in which the factors of success have become

more difficult to identify or imitate (Wade and Hulland 2004).

From a design science perspective, digital artifacts follow a layered-modular

architecture. This structure dissolves product boundaries and allows components to be

loosely coupled and reused while heterogeneous layers can be coupled together

(Adomavicius et al. 2008; Yoo et al. 2010). In turn, the characteristics of digital

products and services have three main implications for firms: They enable the

creation of digital platforms and ecosystems, allow innovation to expand beyond firm

boundaries, and lead to new opportunities for combinatorial innovation (Yoo et al.

2012). For example, different car manufacturers have developed application platforms

or have collaborated with software firms and modified their application platforms.

These platforms provide access to the digital components of smart cars and

incorporate numerous innovative applications developed by independent software developers. Thus by incorporating digital components into their physical products, car manufacturers have been able to draw from sources of innovation that would have been otherwise unavailable. In another example, ridesharing companies have succeeded in using digital technologies to reconfigure physical resources within transportation industries. In turn, these industries have been able to leverage resources from seemingly unrelated industries, such as the food industry, thereby creating an indirect path through a network that enables the flow of digital resources. Our theoretical arguments suggest that this mechanism creates intangible value, enabling firms to generate revenue and profits from their intangible resources.

H1: Firms' intangible value increases with the digital proximity of their products and services.

METHODS

Industry Space

To measure firm-level digital proximity, we begin with the Compustat Historical Segments dataset to construct an industry-level network of complementarities, the "industry space." Figure 1 presents a visualization of the industry space. This data contains the amount of sales reported in different industries for public US firms from 1990 to 2017. We use industry classifications, assuming that at their most detailed level they represent the products and services firms offer.[5] Table 2 provides a closer look at our data, showing that firms generally gain a substantial part of their revenue from industries other than their primary industry. Figure 2 maps the sales reported by automobile manufacturing firms, showing that they participate in

---

[5] The Compustat Historical Segments dataset extends beyond a firm's primary industry classification and shows how each firm's portfolio of sales revenue is segmented among different secondary industry classifications. We use these secondary industry classifications to construct our network dataset.

a wide spectrum of industries, from engine and brake systems to software. Together, Table 2 and Figure 2 help establish the suitability of our data for analyzing cross-industry complementarities.

The industry space network, wherein there is a link between two industries if there is at least one firm that reports sales in both of them, models the complementarities between products and services. Links in this network are created as a result of firms selling in more than one industry; hence, their weights change over time. For example, Bosch, an auto part manufacturer that also sells software solutions such as Business Process Management (BPM) or Internet of Things (IoT) suites; Panasonic, primarily a home appliances manufacturer with sales in both automobile and software industries; and, software firms such as Google and Apple, which also sell car infotainment platforms, all contribute to the creation of links between software and automobile industries. We look at firms' positioning in the industry space to determine their firm-level measures of digital proximity.

Measuring Digital Proximity

The industry space reveals specific paths between firms and digital industries, suggesting opportunities for digitizing products and services. To capture this positioning, we first measure each industry's digital proximity by calculating its average shortest path to software-producing industries. Then, we develop a firm-level measure of digital proximity. We construct this firm-level measure as the average digital proximity of industries a firm operates in, weighted by its relative proportion of sales in each of those industries. Digital proximity captures a firm's diversification into new industries or its change of focus from one group of industries to another. Figure 3 presents a snapshot of our firm-level data. Firms gain high digital proximity scores if they sell directly in digital sectors or sell products with high resource-level

complementarities with digital products. Accordingly, firms with higher digital proximity scores are ahead of their competitors in digitizing their products.

Digital proximity has two unique characteristics, which make it a complement, rather than an alternative, to existing IT investment and innovation measures. First, digital proximity captures the dynamic effects of other firms' activities on the focal firm's performance in the industry space network. Consider Figure 4, which provides a comparison of the trends of IT investment and digital proximity between motion pictures and sound recording industries and petroleum industries. While both industries have comparable levels of IT intensity, their digital proximity scores capture the activities of firms—such as Apple, Amazon, AT&T, and Netflix— involved in digitizing the motion pictures and sound recordings industries. Second, digital proximity captures firms' actual levels of digitization as an outcome-based measure. Figures 4 and 5 reveal how considering only monetary measures of IT investment might be misleading in gauging the extent of firms' digitization. Using such measures might result in assigning inappropriately high levels of digitization to industries such as petroleum or apparel manufacturing relative to more actually digitized industries such as motion pictures and sound recording and motor vehicle manufacturing. In addition, digital proximity reveals firm-level variations in actual levels of digitization (i.e. for broadcasting and telecommunications, securities and financial investments, and insurance industries), which an input-based measure such as IT intensity would not capture. We provide further details regarding the construction of the industry space and the firm-level digital proximity measure in Appendix A.

Data and Variables

For this study, we retrieved data from multiple sources. To create the industry space in which we measure digital proximity, we used data from the Compustat North America Historical Segments database from 1990 to 2017. This data provides firm-year financial measures at the industry-segment level. We then used the Compustat North America database, containing panel data of firms' financial information to construct firm-level and industry-level measures. We then used the Compustat North America Historical Segments data to create weighted averages of measures based on firms' reported sales in each industry. Our analysis focuses on industries that are not conventionally considered to be IT-producing, and thus we have excluded IT firms (e.g. software) from the sample. Ultimately, our dataset includes 126,630 firm-year observations for 14,286 firms, as well as a smaller set of 20,215 firm-year observations for 3,625 firms for which we have complete data for all main model variables. In our robustness checks and supplementary analyses, we used various datasets from SDC Platinum, the Bureau of Labor Statistics, the Bureau of Economic Analysis, and the United States Patent and Trademark Office.

We primarily used Tobin's q, a market-based measure, to calculate firms' intangible value (Bharadwaj et al. 1999; Brynjolfsson et al. 2002; DaDalt et al. 2003). Tobin's q is the ratio of the market value of a company's assets (as measured by the market value of its outstanding stock and debt) divided by the replacement cost of the company's assets (book value). The summary of variables in Table 3 shows the method of calculating Tobin's q. Table 4 shows summary statistics and correlations.

Empirical Models

We used firm-level fixed-effects models to test the relationship between digital proximity and Tobin's q. We controlled for potential confounding effects, such

as size, research and development (R&D) intensity, advertising intensity, intangible assets, market share, industry Herfindahl-Hirschman Index (HHI), as well as the industry average of Tobin's q.

$$\log(q)_{i,t} = \beta_1 * Digital\ Proximity + X_c\beta_c + \sum \beta_t Year_t + u_i + e_{i,t} \quad \text{Eq.}(1)$$

In the above model, *i* and *t* represent each firm and year respectively. X_c represents the aforementioned control variables, u_i represents firm fixed-effects, and $e_{i,t}$ represents errors. The firm fixed-effects model controls for possible differences in relatively stable firm characteristics, including management capabilities and other unobserved characteristics that could introduce confounding effects to the model. In all regression models, we used Huber-White robust standard errors, clustered on the firm identifiers, to account for heteroskedasticity in the errors. The Hausman test—comparing results of fixed-effects and random-effects panel regression estimates—is significant. Hence, the fixed-effects model is preferred for hypothesis testing because, although it is not as efficient, it makes no assumptions about cross-unit correlations between model residuals and regressors. We report the (more consistent) fixed-effects estimates as the main results but also present the (more efficient) random-effects and generalized linear model (GLM)-based estimates. To check for robustness of these estimates, we report ordinary least squares (OLS) regression estimates. Finally, we performed extensive robustness tests to address concerns regarding reverse-causality and unobserved variables.

RESULTS

We find strong support for the hypothesized positive effect of digital proximity on firms' intangible value, as measured by Tobin's q. Table 5 shows the panel regression estimates for the effect of digital proximity on Tobin's q (firm intangible value) using fixed-effect and OLS models. Models (1) and (4) show a

13

positive and significant effect of digital proximity on Tobin's q when controlling for a set of firm- and industry-level variables. Model (1) shows the positive and significant effect ($\beta = 0.191$, $p < 0.01$) of digital proximity on a firm's intangible value in a firm fixed-effects model. Model (4) shows a similar effect ($\beta = 0.161$, $p < 0.05$) in an OLS model. Model (2) shows the positive effect ($\beta = 0.183$, $p < 0.01$) of digital proximity on Tobin's q when including only firm and year fixed-effects. The same effect is tested in Model (5) using an OLS model and is again positive ($\beta = 0.552$) and significant (at $p < 0.01$). Models (3) and (6) show the effect of control variables on Tobin's q in the absence of the main variable; these estimates are comparable to Models (1) and (4) that included the main variable. We ran various post-estimation analyses to ensure the approximate linearity of the relationship between digital proximity and the logarithm of Tobin's q, normal distribution of the error terms with respect to model predictions, insensitivity of the results to outliers, and orthogonality between errors and regressors. We further establish the robustness of our results by re-examining our models with a set of alternative dependent variables and over different sub-samples, reporting the results in Tables 6 and 7.

ROBUSTNESS TESTS

We address potential endogeneity concerns and other empirical concerns through multiple robustness tests and additional empirical analyses.

First, we assess the robustness of the identified relationship between digital proximity and firms' intangible value by constructing a set of instrumental variables, exploiting the network-based nature of the digital proximity measure to do so. We use the average shortest path between the neighbors of a focal industry to every specific digital industry as one group of instrumental variables. These second-order shortest paths from an industry's neighbors to digital industries should have no direct effect on

the intangible value of firms in the focal industry when controlling for all other model covariates; however, they have a direct effect on firms' digital proximity, as they influence the shortest path between focal industries and digital industries. This is because neighbor industries can drag a focal industry closer to digital industries. In our context, any indirect influence of the digital proximity of a focal industry's neighbors on that focal industry's intangible value would occur through the focal industry's own digital proximity via this network-path effect. Accordingly, we calculate the average digital proximity for the set of direct neighbors of each focal industry, which we will henceforth refer to as the *second-order proximity* of the focal industry. Next, we calculate firm-level instrumental variables by constructing the firm-level average of the second-order proximity scores of the industries that firms report sales in, weighted by the amount of sales reported. Table 8 presents the first-stage regression models and Table 9 presents the results of instrumental variable analysis. All models show positive and significant results, providing further support for the main results. In addition, the results of the first-stage regression reported in Table 8 and the various statistical tests reported in Table 9 support the validity of the instruments.

Second, we implement the propensity score matching approach used in prior IS research (Chang and Gurbaxani 2012; Mithas and Krishnan 2009; Oestreicher-Singer and Zalmanson 2013) to further investigate the robustness of our causal identification. We start our analysis by clearly defining the sample, the treatment variable, the outcome variable, and the covariates. We use firm-year observations between 2012 and 2016 for the purpose of matching. In order to form a binary outcome variable, we use the average Tobin's q across firms' industry participation in 2016. We assign a value of 1 for all firms with a Tobin's q score higher than the

15

industry average plus one standard deviation in year 2016, and 0 otherwise. With a similar approach, we develop our binary treatment variable using digital proximity scores and industry averages from 2015.

To construct a list of relevant covariates, including various measures of firms' acquisition and alliancing activities, we use the SDC Platinum database. Strategy and innovation literatures have established mergers, acquisitions, and strategic alliances as important instruments for external innovation (De Man and Duysters 2005; Hagedoorn and Duysters 2002), especially in high-tech industries (Cloodt et al. 2006). IS literature, as well, has explored how mergers, acquisitions, and strategic alliances can enhance firms' ability to digitize their business processes (Trantopoulos et al. 2017) and develop innovative digital products and services (Han et al. 2012; Singh et al. 2015). Our list of covariates includes firms' total number of alliances, joint ventures, software alliances, alliances that include software development, alliances that include marketing services, acquisitions, and software firm acquisitions, as well as the total number of software firm acquisitions in the firm's primary industry between 2011 and 2015.[6] We report the result of the logit regression model predicting the binary form of the treatment variable digital proximity in Table 10. We also perform an additional robustness check with an alternative list of covariates and report these results in Table A2 in Appendix A.[7]

The propensity score matching results, reported in Table 11, provide further support for our hypothesis. We perform various matching techniques and specifications to ensure the robustness of our results within parameters for neighbors and calipers stated in Table 11. The results are consistent across different matching

---

[6] We include the industry-level aggregated measures of acquiring software firms in order to include the industry-level spillovers of other firms' digitization through acquisitions (Cheng and Nault 2007; Tambe and Hitt 2013).
[7] In this robustness test, we repeated the procedure for an alternative group of covariates, such as sales efficiency, R&D intensity, and intangible assets that are associated with firms' level of digitization, which we had access to from the Compustat dataset (Chang and Gurbaxani 2012; Im et al. 2001; Kleis et al. 2012).

techniques (Kernel matching in Column (1) and nearest neighbor matching in Columns (2) to (4)).

One strong assumption underlying our matching method is that we can assess firms' digital proximity using the set of identified covariates; however, identifying the complete set of factors affecting digital proximity and providing an exact prediction of the treatment variable is not feasible. In light of the possibility of biased estimates from unobserved factors (Caliendo and Kopeinig 2008), we assess the sensitivity of our results to the effects of various unobserved factors. Accordingly, we examine the probability of classifying a firm in the treatment group or the control group based on unobserved factors (Rosenbaum and Rubin 1983) and report the results, critical Rosenbaum's Gamma values, in Table 11.[8] These values fall within an acceptable range (Mithas and Krishnan 2009; Oestreicher-Singer and Zalmanson 2013), between 1.8 to 4.2. Thus, unless we can argue that the unobserved factors can change the odds of having a high level of digital proximity by at least 80 percent, we can rely on the previously generated estimates. Given that we capture an acceptable level of variation for our set of observed factors, we can conclude that our results are robust to unobserved factors. Table 12 presents the descriptive statistics of covariances before and after matching.

Further, we perform additional falsification tests to ensure the identified effect is not attributed to chance and common industry trends, reporting these results in Table 13. Following recent research (Burtch and Chan 2019; Lee and Hosanagar 2019), we use the shuffling technique in three separate tests to consider whether the

---

[8] Gamma values are calculated based on Rosenbaum bounds sensitivity analysis (Rosenbaum 2002). In our context, this sensitivity analysis provides an upper bound for the effect of unobserved factors to nullify the effect of digital proximity on firms' Tobin's q. We construct the odds ratio Gamma (log odds of differential assignment to treatment due to unobserved factors), which measures the extent to which firms with the same observed attributes differ in their levels of digital proximity. If Gamma equals 1.8 for 2 firms with similar observed characteristics, their propensity to have a high digital proximity score is different by 80 percent, or a factor of 1.8.

identified relationship arises spuriously. First, we randomly assign digital proximity scores without considering the industry and year panel variables. Then, we randomly assign these scores within the same industries and, finally, within the same years in the same industries. For each test, we conduct random shuffling 100 times, re-estimating the coefficient for each permutation and reporting the mean and standard error of these 100 coefficient estimates for digital proximity. Lack of statistical significance in all of these models shows that our results are independent of industrial trends and environmental changes. For further details on this procedure, see Good (2005).

Finally, we assess the face validity of the digital proximity measure by comparing it to three measures: (1) similarity to digital knowledge, (2) knowledge complexity, and (3) digital innovations. We report this analysis in Appendix B. We perform additional tests and discuss the robustness of our results to multicollinearity, unobserved heterogeneities, and reverse causality in Appendix C.

DISCUSSION

Research Implications

Inspired by the literature examining the effect of proximity on the innovation output of firms and economies (Green 2018; Hidalgo et al. 2007; Klepper 2010), we introduce the digital proximity construct and its measurement, which captures the closeness of a firm's products and services to digital industries. We examine the relationship between digital proximity and intangible value over the past several decades by constructing an industry space that shows how firms' proximity to digital industries evolves over time.

Introducing digital proximity, as a theoretical construct, has two specific theoretical contributions to the field of IS. First, this work explores IT beyond its

18

being a monetary investment with productivity effects, and keys in on the intangible value of digitization of products and services as a form of strategic positioning within the economic network linking all industries. With digital technologies becoming increasingly pervasive in a growing number of products and services previously considered to be non-digital, our results highlight digital innovation as an inherent driver—rather than an appendage of—firms' business strategy. This implication aligns with prior literature that recognizes IT as an inherent component of firms' business strategy (Bharadwaj et al. 2013). It also aligns with the contemporary reality of digital transformation (Yoo et al. 2012). By extending research beyond the factors that help firms establish superior IT capabilities, the current investigation motivates further studies on the question of how firms can gain high levels of digitization. Moving forward, more research is needed to better understand the firm-level and environmental factors that could impact a firm's ability to translate digitization success into intangible value and a sustainable competitive advantage.

Second, we bring a novel approach to a central and persistent research question about the role of digital resources and capabilities in generating firm value in a hyper-competitive environment (Lee et al. 2010; Rai and Tang 2013; Tanriverdi et al. 2010). Our study achieves this by introducing a novel network-based construct, which considers a competitive landscape in defining firms' success in digitizing their products and services. Because the economy has moved into an age of distributed digital innovation, we believe that the IS literature will benefit from a network-based construct such as what we present in this work, defining firms' success in digitizing their products and services in relation to the dynamic underlying complementarity networks upon which firms operate. Digital proximity entails the enhancement or deterioration of firms' competitive position, even when this movement is caused by

the activities of other firms. As a network-based construct, digital proximity considers the macro patterns of change in the overall economic landscape in the movement of firms toward digitization. This approach in analyzing digital innovation empowers researchers to explore digital innovations as generative and interconnected, rather than isolated (Yoo et al. 2012; Yoo et al. 2010).

Our novel approach in measuring digitization contributes substantively to the business value of IT literature. First, we measure digital proximity as an outcome-based measure that captures digitization. That is, our measure entails a wide range of factors, such as IT investments and patenting activities, and focuses on how firms translate these factors successfully into digital products and services. In this way, digital proximity reveals another IT capability, one that is not captured by traditional measures. Second, the network-based measure of digital proximity includes firms' positioning in a competitive landscape when assessing their levels of digitization. Traditional measures of digitization, however, focus only on single firms' IT investments; in other words, these measures treat firms as isolated entities. Using our measure, a firm's positioning can improve if other firms in its industry create new complementarities between industry resources and digital resources. Such inter-firm effects are not captured in traditional IT investment measures. Third, our measurement procedure overcomes a persistent limitation affecting the business value of IT research, which stems from the field's reliance on survey data and the unavailability of many firms' IT investment data. Our approach in measuring digitization has allowed us to base our findings on 20,215 observations from 3,625 unique firms over a time period of almost three decades.

Managerial Implications

Our findings have important practical implications, as well. First, our results highlight the importance of considering a wider competitive landscape in today's digital economy. We show how the activities of firms in other industries might create or undermine firms' paths to digitization. Therefore, in addition to their immediate competitors and partners, firms need strategic insight into other industries that might emerge to create connecting paths to digital industries over time. We also develop the industry space, which maps firms' competitive environment, and managers could use this network to analyze the relative digital positioning of their firm against competitors and identify potential areas from which new competition may arise. In turn, this work could guide managers when choosing among paths of different complementary technologies.

Second, our results introduce a novel practical approach for measuring digitization within a wider economic landscape. In this way, our results provide a response to a call by economics researchers (Greenstein et al. 2013) regarding the need for new measures that capture digitization at a macroeconomic level. Our firm-level and industry-level measures of digital proximity equip policy makers with a new tool to measure overall movement toward digitization. For example, an overall average industry-level digital proximity score can measure the digitization of an economy, the same way that GDP captures its production.

REFERENCES

Adomavicius, G., Bockstedt, J. C., Gupta, A., and Kauffman, R. J. 2008. "Making Sense of Technology Trends in the Information Technology Landscape: A Design Science Approach," *MIS Quarterly* (32:4), pp. 779-809 (doi: 10.2307/25148872).

Amit, R., and Schoemaker, P. J. 1993. "Strategic Assets and Organizational Rent," *Strategic Management Journal* (14:1), pp. 33-46 (doi: 10.1002/smj.4250140105).

Aral, S., and Weill, P. 2007. "IT Assets, Organizational Capabilities, and Firm Performance: How Resource Allocations and Organizational Differences

Explain Performance Variation," *Organization Science* (18:5), pp. 763-780 (doi: 10.1287/orsc.1070.0306).

Baldwin, C., and Von Hippel, E. 2011. "Modeling a Paradigm Shift: From Producer Innovation to User and Open Collaborative Innovation," *Organization Science* (22:6), pp. 1399-1417. (doi: 10.1287/orsc.1100.0618).

Barney, J. 1991. "Firm Resources and Sustained Competitive Advantage," *Journal of Management* (17:1), pp. 99-120.

Barrett, M., Davidson, E., Prabhu, J., and Vargo, S. L. 2015. "Service Innovation in the Digital Age: Key Contributions and Future Directions," *MIS Quarterly* (39:1), pp. 135-154.

Bharadwaj, A., Bharadwaj, S., and Konsynski, B. 1999. "Information Technology Effects on Firm Performance as Measured by Tobin's q," *Management Science* (45:7), pp. 1008-1024.

Bharadwaj, A., El Sawy, O. A., Pavlou, P. A., and Venkatraman, N. V. 2013. "Digital Business Strategy: Toward a Next Generation of Insights," *MIS Quarterly* (37:2), pp. 471-482.

Bhatt, G. D., and Grover, V. 2005. "Types of Information Technology Capabilities and Their Role in Competitive Advantage: An Empirical Study," *Journal of Management Information Systems* (22:2), pp. 253-277 (doi: 10.1080/07421222.2005.11045844).

Boschma, R., Balland, P.-A., and Kogler, D. F. 2015. "Relatedness and Technological Change in Cities: The Rise and Fall of Technological Knowledge in US Metropolitan Areas from 1981 to 2010," *Industrial and Corporate Change* (24:1), pp. 223-25 (doi: 10.1093/icc/dtu012).

Branstetter, L. G., Drev, M., and Kwon, N. 2019. "Get with the Program: Software-driven Innovation in Traditional Manufacturing," *Management Science* (65:2), pp. 541-558).

Brynjolfsson, E., and Hitt, L. 1996. "Paradox Lost? Firm-level Evidence on the Returns to Information Systems Spending," *Management Science* (42:4), pp. 541-558.

Brynjolfsson, E., and Hitt, L. M. 2000. "Beyond Computation: Information Technology, Organizational Transformation and Business Performance," *The Journal of Economic Perspective* (14:4), pp. 23-48.

Brynjolfsson, E., Hitt, L. M., and Yang, S. 2002. "Intangible Assets: Computers and Organizational Capital," *Brookings Papers on Economic Activity* (2002:1), pp. 137-181.

Burtch, G., and Chan, J. 2019. "Investigating the Relationship Between Medical Crowdfunding and Personal Bankruptcy in the United States: Evidence of a Digital Divide," *MIS Quarterly* (43:1), pp. 237-262 (doi: 10.25300/MISQ/2019/14569).

Caliendo, M., and Kopeinig, S., 2008. "Some Practical Guidance for the Implementation of Propensity Score Matching," *Journal of economic surveys* (22:1), pp. 31-72.

Chae, H.-C., Koh, C. E., and Prybutok, V. R. 2014. "Information Technology Capability and Firm Performance: Contradictory Findings and Their Possible Causes," *MIS Quarterly* (38:1), pp. 305-326.

Chaney, T. 2014. "The Network Structure of International Trade," *American Economic Review* (104:11), pp. 3600-3634 (doi: 10.1257/aer.104.11.3600).

Chang, Y. B., and Gurbaxani, V. 2012. "Information Technology Outsourcing, Knowledge Transfer, and Firm Productivity: An Empirical Analysis," *MIS Quarterly* (36:4), pp. 1043-1063 (doi: 10.2307/41703497).

Cheng, Z., and Nault, B. R. 2007. "Industry Level Supplier-driven IT Spillovers," *Management Science* (53:8), pp. 1199-1216 (doi: 10.1287/mnsc.1060.0657).

Cloodt, M., Hagedoorn, J., and Van Kranenburg, H. 2006. "Mergers and Acquisitions: Their Effect on the Innovative Performance of Companies in High-tech Industries," *Research Policy* (35:5), pp. 642-654 (doi: 10.1016/j.respol.2006.02.007).

DaDalt, P. J., Donaldson, J. R., and Garner, J. L. 2003. "Will Any q Do?" *Journal of Financial Research* (26:4), pp. 535-551 (doi: 10.1111/1475-6803.00073).

De Man, A.-P., and Duysters, G. 2005. "Collaboration and Innovation: A Review of the Effects of Mergers, Acquisitions and Alliances on Innovation," *Technovation* (25:12), pp. 1377-1387 (doi: 10.1016/j.technovation.2004.07.021).

Dyer, J. H., and Singh, H. 1998. "The Relational View: Cooperative Strategy and Sources of Interorgal Competitive Advantage," *The Academy of Management Review* (23:4), pp. 660-679 (doi: 10.5465/amr.1998.1255632).

Eisenhardt, K. M., and Martin, J. A. 2000. "Dynamic Capabilities: What Are They?" *Strategic Management Journal* (21:10/11), pp. 1105-1121.

Forman, C., Goldfarb, A., and Greenstein, S. 2016. "Agglomeration of Invention in the Bay Area: Not Just ICT," *American Economic Review* (106:5), pp. 146-151.

Gawer, A. 2014. "Bridging Differing Perspectives on Technological Platforms: Toward an Integrative Framework," *Research Policy* (43:7), pp. 1239-1249 (doi: 10.1016/j.respol.2014.03.006).

Good, P. 2005. *Permutation, Parametric, and Bootstrap Tests of Hypotheses*, New York, NY: Springer.

Green, D. D. 2018. "Fueling the Gig Economy: A Case Study Evaluation of Upwork.com," *Management and Economics Research Journal* (4:2018), pp. 104-112 (doi: 10.18639/MERJ.2018.04.523634).

Greenstein, S., Lerner, J., and Stern, S. 2013. "Digitization, Innovation, and Copyright: What is the Agenda?" *Strategic Organization* (11:1), pp. 110-121.

Guevara, M. R., Hartmann, D., Aristarán, M., Mendoza, M., and Hidalgo, C. A. 2016. "The Research Space: Using the Career Paths of Scholars to Predict the Evolution of the Research Output of Individuals, Institutions, and Nations," *arXiv preprint arXiv:1602.08409*.

Hacklin, F., Battistini, B., and Von Krogh, G. 2013. "Strategic Choices in Converging Industries," *MIT Sloan Management Review* (55:1), pp. 65-74.

Hagedoorn, J., and Duysters, G. 2002. "External Sources of Innovative Capabilities: The Preferences for Strategic Alliances or Mergers and Acquisitions," *Journal of Management Studies* (39:2), pp. 167-188.

Han, K., Oh, W., Im, K. S., Chang, R. M., Oh, H., and Pinsonneault, A. 2012. "Value Cocreation and Wealth Spillover in Open Innovation Alliances," *MIS Quarterly* (36:1), pp. 291-315.

Hartmann, D., Guevara, M. R., Jara-Figueroa, C., Aristarán, M., and Hidalgo, C. A. 2017. "Linking Economic Complexity, Institutions, and Income Inequality," *World Development* (93), pp. 75-93 (doi: 10.1016/j.worlddev.2016.12.020).

Hausmann, R., and Hidalgo, C. A. 2011. "The Network Structure of Economic Output," *Journal of Economic Growth* (16:4), pp. 309-342 (doi: 10.1007/s10997-011-9071-4).

Hausmann, R., Hwang, J., and Rodrik, D. 2007. "What You Export Matters," *Journal of Economic Growth* (12:1), pp. 1-25 (doi: 10.1007/s10887-006-9009-4).

Henfridsson, O., and Bygstad, B. 2013. "The Generative Mechanisms of Digital Infrastructure Evolution," *MIS Quarterly* (37:3), pp. 907-931.

Hidalgo, C. 2015. *Why Information Grows: The Evolution of Order, from Atoms to Economies*, New York, .NY: Basic Books.

Hidalgo, C., and Hausmann, R. 2009. "The Building Blocks of Economic Complexity," *Proceedings of the National Academy of Sciences* (106:26), pp. 10570-10575 (doi: 10.1073/pnas.0900943106).

Hidalgo, C. A., Klinger, B., Barabási, A.-L., and Hausmann, R. 2007. "The Product Space Conditions the Development of Nations," *Science* (317:5837), pp. 482-487 (doi: 10.1126/science.1144581).

Im, K. S., Dow, K. E., and Grover, V. 2001. "Research Report: A Reexamination of IT Investment and the Market Value of the Firm—An Event Study Methodology," *Information Systems Research* (12:1), pp. 103-117.

Kleis, L., Chwelos, P., Ramirez, R. V., and Cockburn, I. 2012. "Information Technology and Intangible Output: The Impact of IT Investment on Innovation Productivity," *Information Systems Research* (23:1), pp. 42-59 (doi: 10.1287/isre.1100.0338)/

Klepper, S. 2010. "The Origin and Growth of Industry Clusters: The Making of Silicon Valley and Detroit," *Journal of Urban Economics* (67:1), pp. 15-32.

Kohli, R., and Grover, V. 2008. "Business Value of IT: An Essay on Expanding Research Directions to Keep up with the Times," *Journal of the Association for Information Systems* (9:1), pp. 23-39).

Lee, C. H., Venkatraman, N., Tanriverdi, H., and Iyer, B. 2010. "Complementarity-based Hypercompetition in the Software Industry: Theory and Empirical Test, 1990–2002," *Strategic Management Journal* (31:13), pp. 1431-1456 (doi: 10.1002/smj.895).

Lee, D., and Hosanagar, K. 2019. "How Do Recommender Systems Affect Sales Diversity? A Cross-Category Investigation via Randomized Field Experiment," *Information Systems Research* (30:1), pp. 239-259.

Maznevski, M. L., and Chudoba, K. M. 2000. "Bridging Space over Time: Global Virtual Team Dynamics and Effectiveness," *Organization Science* (11:5), pp. 473-492.

McGrath, R. G., Tsai, M.-H., Venkataraman, S., and MacMillan, I. C. 1996. "Innovation, Competitive Advantage and Rent: A Model and Test," *Management Science* (42:3), pp. 389-403.

Melville, N., Kraemer, K., and Gurbaxani, V. 2004. "Review: Information Technology and Organizational Performance: An Integrative Model of IT Business Value," *MIS Quarterly* (28:2), pp. 283-322 (doi: 10.2307/25148636).

Mithas, S., and Krishnan, M. S. 2009. "From Association to Causation via a Potential Outcomes Approach," *Information Systems Research* (20:2), pp. 295-313.

Mithas, S., Tafti, A., Bardhan, I., and Goh, J. M. 2012. "Information Technology and Firm Profitability: Mechanisms and Empirical Evidence," *MIS Quarterly* (36:1), pp. 205-224.

Mocker, M., Weill, P., and Woerner, S. L. 2014. "Revisiting Complexity in the Digital Age," *MIT Sloan Management Review* (55:4), pp. 73-82.

Nan, N., and Tanriverdi, H. 2017. "Unifying the Role of IT in Hyperturbulence and Competitive Advantage via a Multilevel Perspective of IS Strategy," *MIS Quarterly* (41:3), pp. 937-958.

Nandkumar, A., Mani, D., and Bharadwaj, A. 2018. "Market Returns to Digital Innovations: A Group Based Trajectory Approach," *Academy of Management Proceedings* (2018:1) (doi: 10.5465/AMBPP.2018.15241abstract).

Neffke, F., Henning, M., and Boschma, R. 2011. "How Do Regions Diversify over Time? Industry Relatedness and the Development of New Growth Paths in Regions," *Economic Geography* (87:3), pp. 237-265 (doi: 10.1111/j.1944-8287.2011.01121.x).

Novales, A., Mocker, M., and Simonovich, D. 2016. "IT-enriched "Digitized" Products: Building Blocks and Challenges," *Proceedings of the Americas Conference on Information Systems* (22:2), pp. 1595-1604.

Oestreicher-Singer, G., and Zalmanson, L. 2013. "Content or Community? A Digital Business Strategy for Content Providers in the Social Age," *MIS Quarterly* (37:2), pp. 591-616.

Parmar, R., Mackenzie, I., Cohn, D., and Gann, D. 2014. "The New Patterns of Innovation," *Harvard Business Review* (92:1/2), pp. 86-95.

Porter, M. E., and Heppelmann, J. E. 2015. "How Smart, Connected Products Are Transforming Competition," *Harvard Business Review* (93:10), pp. 96-116.

Powell, T. C., and Dent-Micallef, A. 1997. "Information Technology as Competitive Advantage: The Role of Human, Business, and Technology Resources," *Strategic Management Journal* (18:5), pp. 375-405.

Rai, A., Patnayakuni, R., and Patnayakuni, N. 1997. "Technology Investment and Business Performance," *Communications of the ACM* (40:7), pp. 89-97

(doi: 10.1145/256175.256191).

Rai, A., and Tang, X. 2013. "Research Commentary—Information Technology-enabled Business Models: A Conceptual Framework and a Coevolution Perspective for Future Research," *Information Systems Research* (25:1), pp. 1-14.

Ray, G., Barney, J. B., and Muhanna, W. A. 2004. "Capabilities, Business Processes, and Competitive Advantage: Choosing the Dependent Variable in Empirical Tests of the Resource-based View," *Strategic Management Journal* (25:1), pp. 23-37.

Rodrik, D. 2006. "What's So Special about China's Exports?" *China & World Economy* (14:5), pp. 1-19.

Rosenbaum, P. 2002. "Observational Studies," 2nd ed. Springer, New York.

Rosenbaum, P. R., and Rubin, D. B. 1983. "Assessing Sensitivity to an Unobserved Binary Covariate in an Observational Study with Binary Outcome," *Journal of the Royal Statistical Society: Series B (Methodological)*, (45:2), pp. 212-218.

Singh, R., Mathiassen, L., and Mishra, A. 2015. "Organizational Path Constitution in Technological Innovation: Evidence from Rural Telehealth," *MIS Quarterly* (39:3), pp. 643-666.

Tambe, P., and Hitt, L. M. 2013. "Measuring Information Technology Spillovers," *Information Systems Research* (25:1), pp. 53-71.

Tanriverdi, H., Rai, A., and Venkatraman, N. 2010. "Research Commentary—Reframing the Dominant Quests of Information Systems Strategy Research for Complex Adaptive Business Systems," *Information Systems Research* (21:4), pp. 822-834.

Teece, D. J., Pisano, G., and Shuen, A. 1997. "Dynamic Capabilities and Strategic Management," *Strategic Management Journal* (18:7), pp. 509-533.

Trantopoulos, K., von Krogh, G., Wallin, M. W., and Woerter, M. 2017. "External Knowledge and Information Technology: Implications for Process Innovation Performance," *MIS Quarterly* (41:1), pp. 287-300.

Wade, M., and Hulland, J. 2004. "The Resource-based View and Information Systems Research: Review, Extension, and Suggestions for Future Research," *MIS Quarterly* (28:1), pp. 107-142 (doi: 10.2307/25148626).

Yoo, Y., Boland Jr, R. J., Lyytinen, K., and Majchrzak, A. 2012. "Organizing for Innovation in the Digitized World." *Organization Science* (23:5), pp. 1398-1408 (doi: 10.1287/orsc.1120.0771).

Yoo, Y., Henfridsson, O., and Lyytinen, K. 2010. "Research Commentary—The New Organizing Logic of Digital Innovation: An Agenda for Information Systems Research," *Information Systems Research* (21:4), pp. 724-735.

Figure 1: Industry space for the time period 2010–2014, visualized using a force expansion algorithm. Nodes represent industry classifications at the 6-digit NAICS level, and they are colored according to the 2-digit NAICS level. Node weights are proportional to their degree, and the weights of the edges are set to be equal. The IT sector is shown in yellow, so the network can reveal the relatedness of various sub-industries. Manufacturing (in red) features a large amount of dispersion with respect to its proximity to IT. Sub-industries such as semiconductor manufacturing, surgical and medical equipment manufacturing, and aerospace equipment manufacturing are relatively close to IT, whereas chemical manufacturing, fabricated metal manufacturing, and hardware manufacturing are relatively far from IT. The manufacturing cluster is surrounded by information industries (yellow, bottom right); finance and insurance (dark blue); healthcare and social services (dark green); mining, oil and gas; transportation; and, agriculture. In the overall industry space, several information-intensive industries are, as expected, particularly close to IT: electronic shopping; professional, scientific, and technological services; art, entertainment, and recreation; finance; and, high-tech manufacturing industries. Industries such as oil and gas, agriculture, mining, utilities, dairy production, and textiles are located far from IT.

27
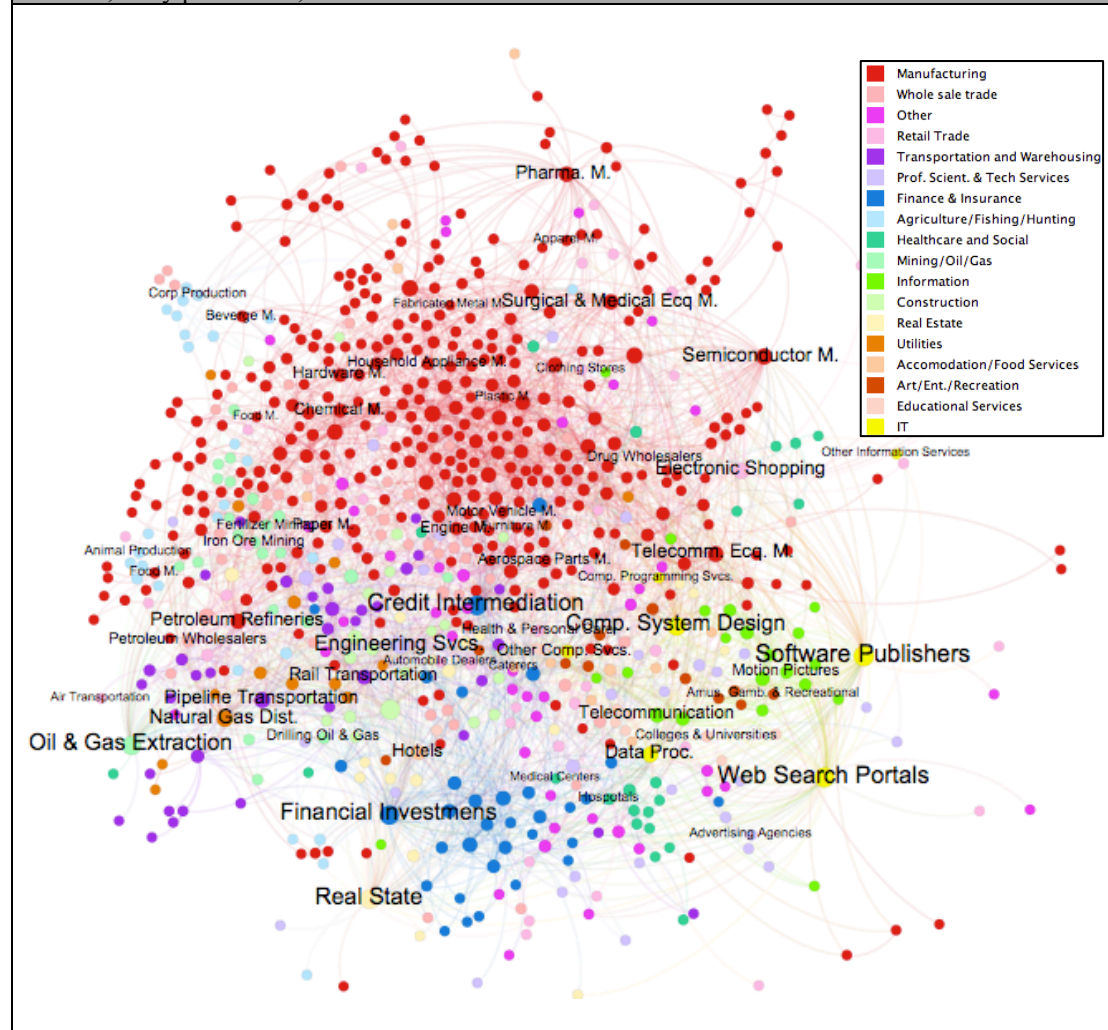
Figure 2: Map of sales reported by all firms in the automobile industries over the industry space network with selective labels (top). Red in this network represents an industry that a firm originally classified in the motor vehicle manufacturing industries (NAICS 361) reports sales in. Blue nodes are all other industries. This figure shows how firms in automobile industries have sales in a wide spectrum of industries, including engine and brake systems, navigation and control instruments, audio and video equipment, semiconductor and electronic equipment, computer system design, and software industries. Links in the industry space are created as a result of firms selling in more than one industry (bottom). Specifically, Google is a software firm that is developing its own self-driving vehicle, Bosch is an auto part manufacturer that also sells software solutions, and Panasonic is a home appliances manufacturer that also sells automobile parts and cloud solutions.
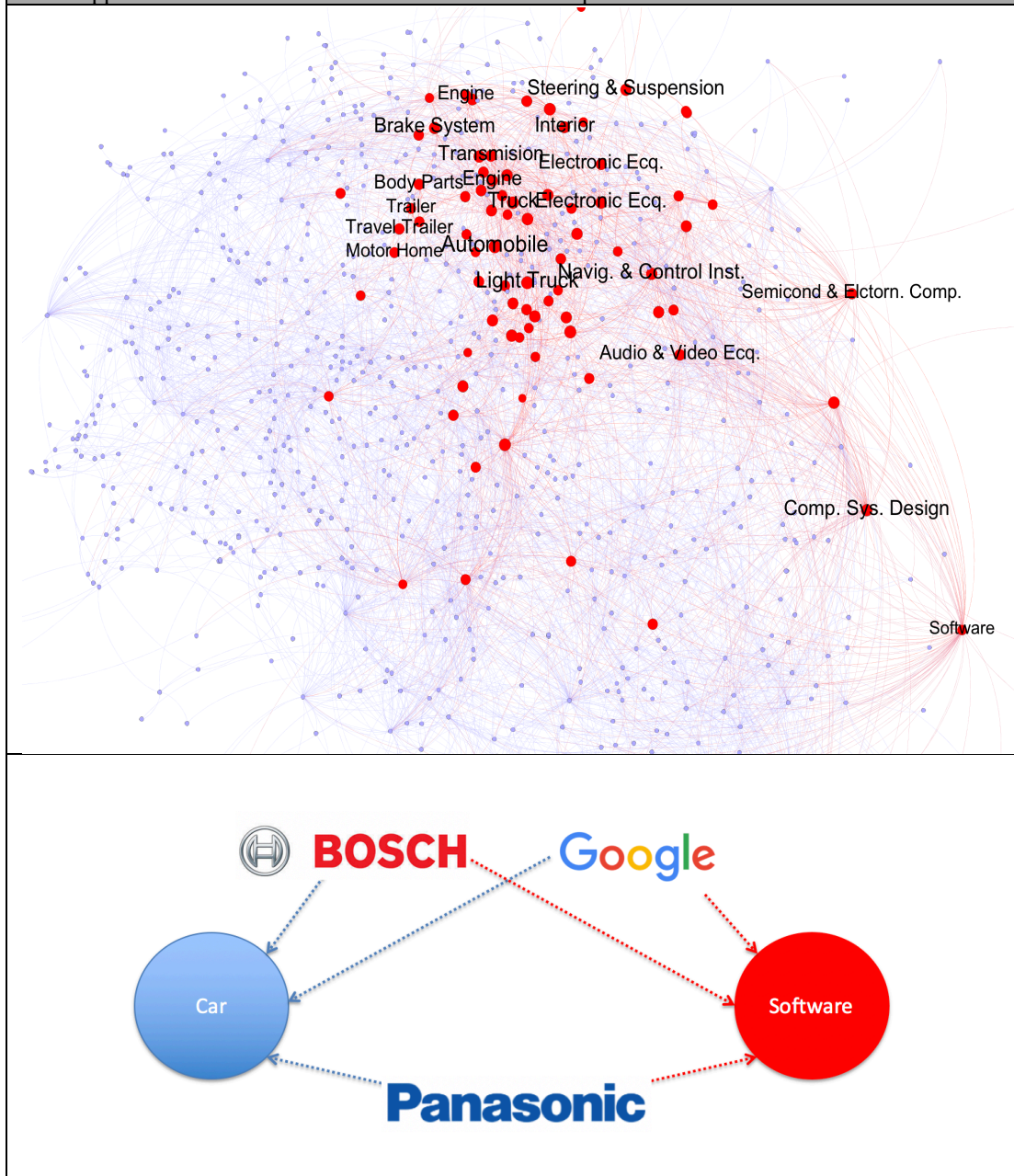
Figure 3: Snapshot of firms in the wholesale trade industries (red) in 2015 and 2016, sorted by their digital proximity scores (from low at the top to high at the bottom) and the industries they report sales in (blue), sorted by their digital proximity scores (from low at the top to high at the bottom). Firm node size is proportional to Tobin's q, and industry node size is proportional to digital proximity scores. Two takeaways enhancing the validity of the digital proximity construct: (1) while communication, finance, and aircraft manufacturing industries have higher levels of digital proximity, book and newspaper wholesalers, trust manufacturers, food, and construction industries have lower digital proximity scores, and (2) firms such as usell.com, an online platform for trading used electronic devices; Bsquare, a provider of industrial internet of things (IoT) suits; and, Inpixon, a provider of indoor positioning analytics security and intelligence products, have the highest digital proximity scores. Meanwhile, firms such as Now Inc, a distributor of supplies, and Boise Cascade, a wood products distributor, have the lowest digital proximity scores.
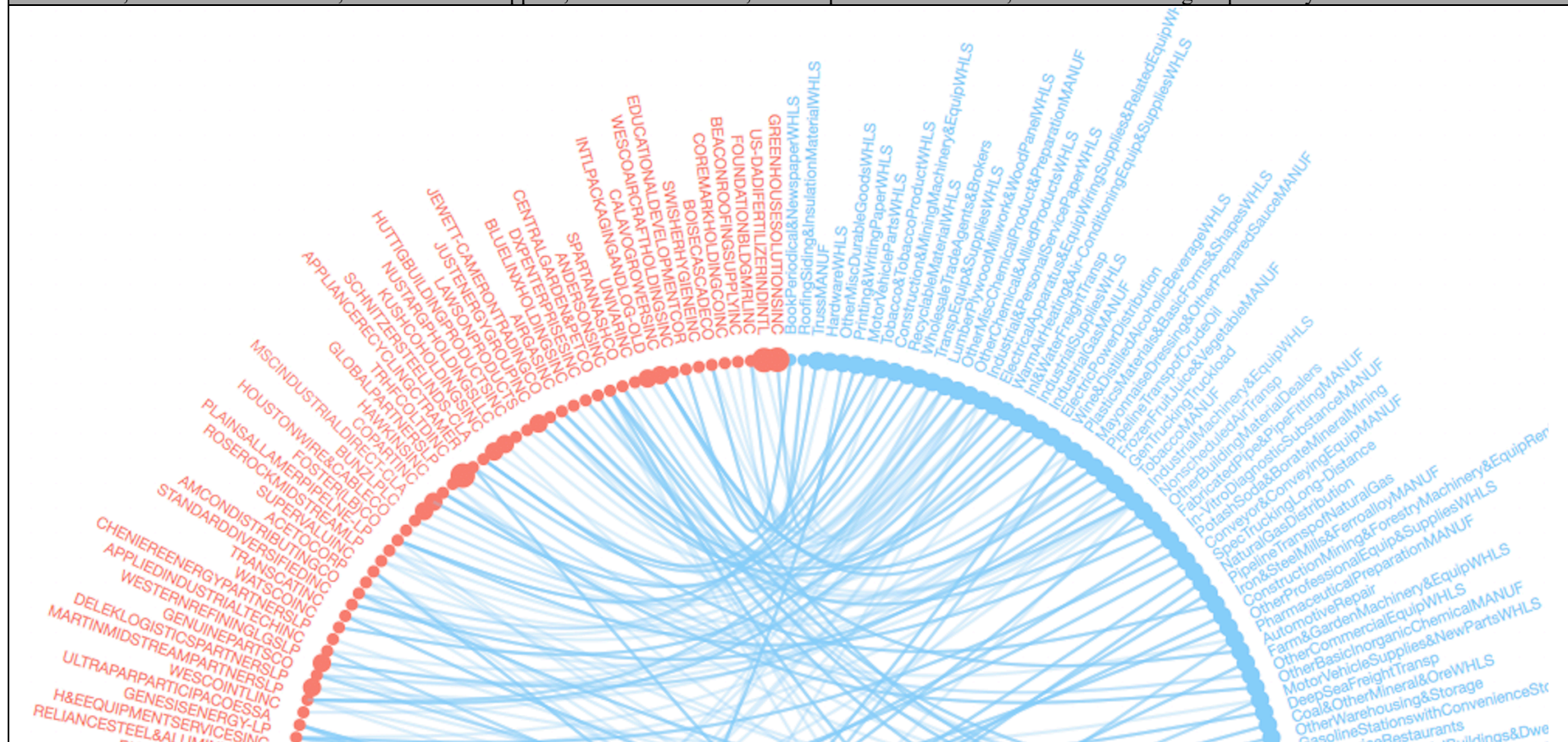
Figure 3: (Continued)
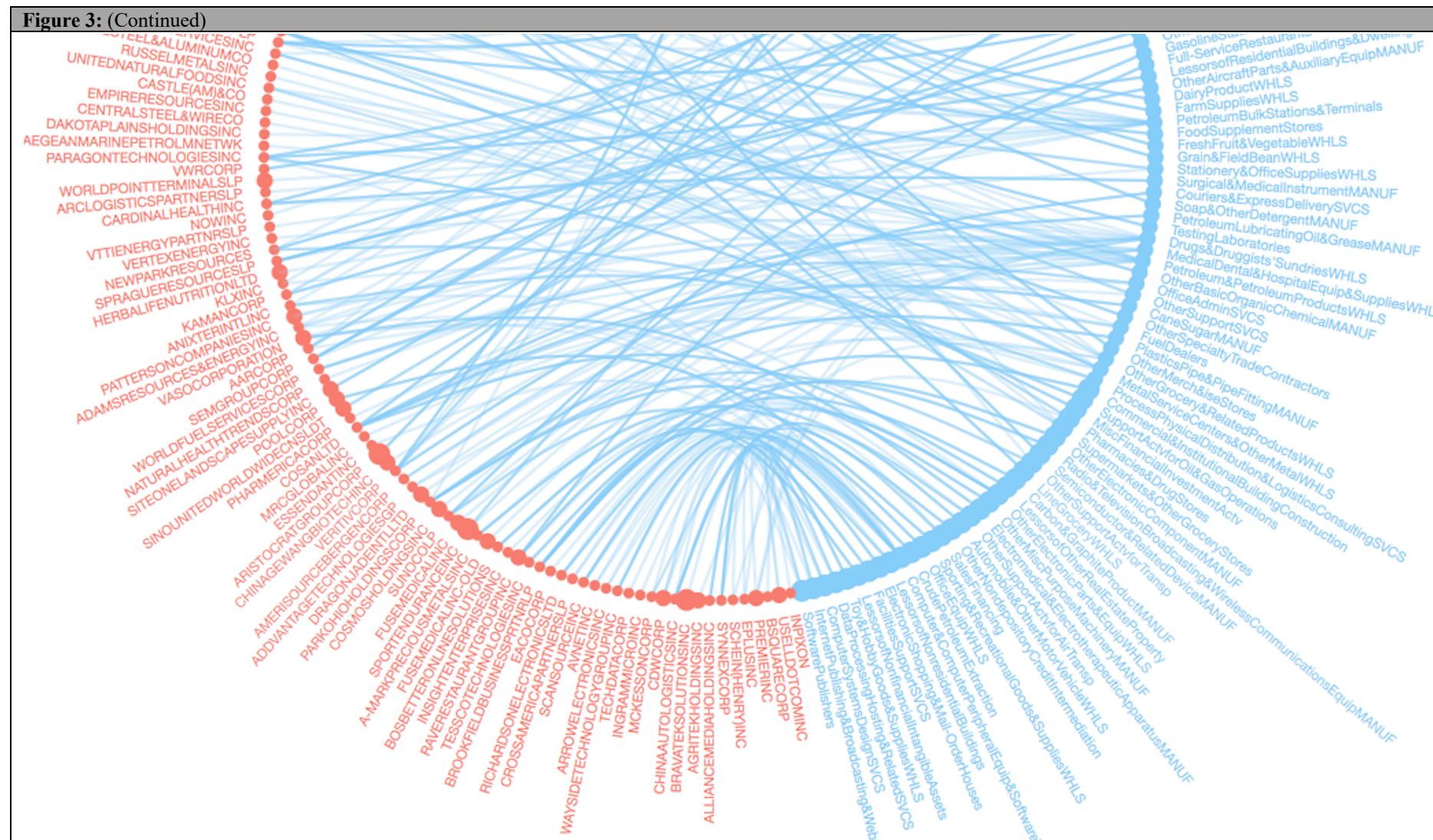
Figure 4: Comparison of digital proximity and IT investments for motion pictures and sound recording industries (NAICS 512) and petroleum industries (NAICS 324) over time (top). Digital proximity co-evolves with the activities of other firms in the industry. These variations, while reflected in the digital proximity measure, are not captured by traditional constructs, such as monetary IT investments. IT investment data is extracted from the Bureau of Economic Analysis (BEA) industry-level data on capital investments. The IT investment variable is weighted by industry-level total capital investments. All IT investment and digital proximity values are mean centered and standardized.
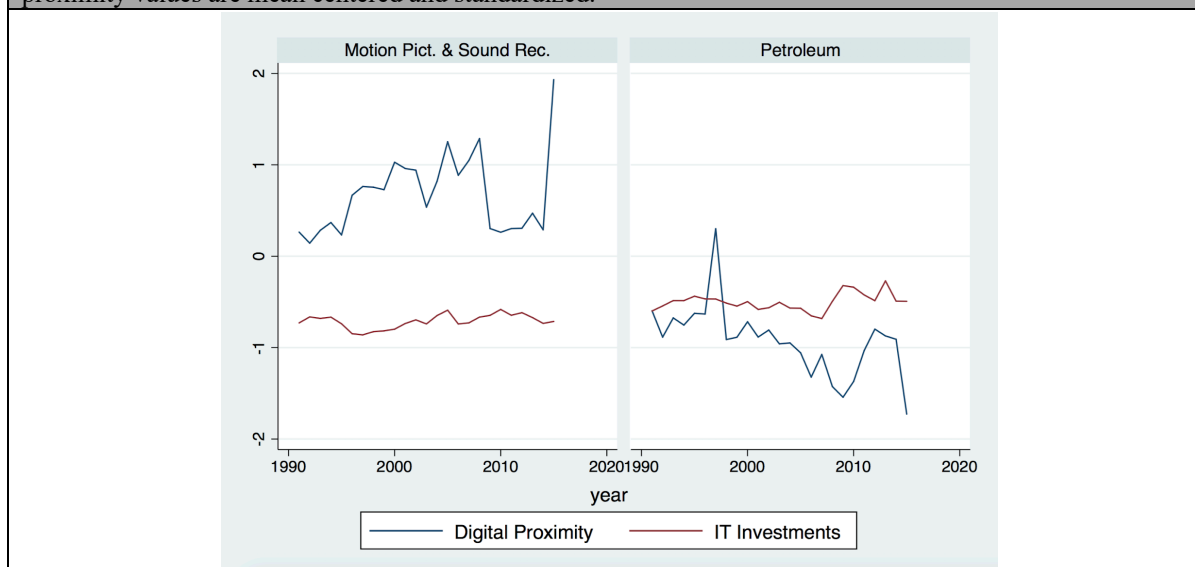
Figure 5: Relationship between industry-level (3-digit NAICS) IT investments and digital proximity. Digital proximity captures firms' actual levels of digitization. While the two measures are correlated, meaningful variation is not captured by monetary IT investments. For example, apparel manufacturing industries have a higher level of IT investments compared to motion pictures and sound recording industries, while the latter have a higher value of digital proximity. Further, broadcasting and telecommunications, securities and financial investments, and insurance firms have similar levels of IT investment; however, their digital proximity scores clearly show their different levels of digitization. In the below graph, IT-producing (e.g. software) industries are excluded. IT investment data is extracted from the Bureau of Economic Analysis (BEA) industry-level data on capital investments. The IT investment variable is weighted by industry-level total capital investments. All IT investment and digital proximity values are mean centered and standardized.

31
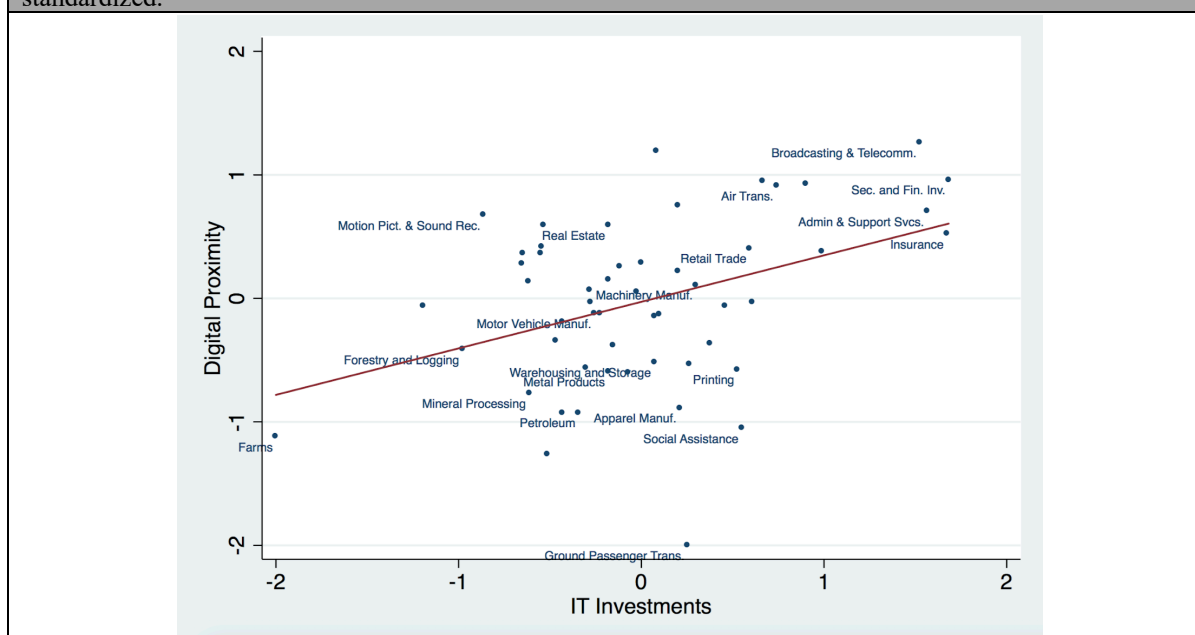
Table 1: Summary of literature on the performance effects of digitization. The below table is not a comprehensive literature review and lists only a few studies for each theoretical construct.					
Literature	**Theoretical Construct**	**Construct Definition**	**Measure**	**Outcome-based Measure**	**Network-based Measure**
Aral and Weill 2007; Brynjolfsson and Hitt 1996; Mithas et al. 2012; Rai et al. 1997	IT investments	Investments made in IT systems and applications.	Annual IT budget: Annual IT budget per employee, or as a percentage of sales.	No	No
Bharadwaj 2000; Chae et al. 2014; Santhanam and Hartono 2003	IT capabilities	Ability to combine individual IT resources into superior applications.	IT leaders: If a firm is famous for being a technology leader or developing and implementing innovative or strategic applications.	No	No
Branstetter et al. 2018; Nan and Tanriverdi 2017	Digital innovation	Developing digital modules for non-digital products and services.	Software patents as a share of total patents in non-software industries (Branstetter et al. 2018).	No	No
			Agent-based modeling simulation (Nan and Tanriverdi 2017).	N/A	N/A
This paper	Digital proximity	Digitizing products and services by accumulating the knowledge and know-how required for their development.	The proximity of firms' sales portfolios to those of digital firms across a network of complementarities.	Yes	Yes

Table 2: Descriptive statistics for the Compustat Segments dataset and our final sample, specifically (1) portion of sales reported by firm-year observations in industries other than the primary industry, (2) average number of sub-industries that firm-year observations report sales in, (3) ratio of firm-year observations that report sales in more than one industry over the total number of firms in each industry, and (4) portion of firms that report sales in more than one industry. Utility industries (NAICS 22) have the lowest (0.57) average ratio of sales in other industries to sales in the primary industry; food, textile, and apparel manufacturing industries have the highest (0.87). There is a large variation in the average number of sub-industries that firms report sales in. Industries such as agriculture, fishing, and forestry (NAICS 11); utility (NAICS 22); and, construction (NAICS 23) have more diversified firm-year observations, while retail trade industries (NAICS codes 44 and 45) and real estate industries (NAICS 53) have less diversified firm-year observations. Finally, except for oil and gas industries (NAICS 21) and other services industries (NAICS 81), at least 45 percent of the firms in all industry groups report sales in more than one industry.

		Compustat Segments Data				
NAICS 2-Digit Code	Ratio of Sales in Other Inds. over Sales in Primary Ind.	Average Number of Sub-inds. for Firms-years	Firms-years with Sale in Other Inds. to Total Firms	Portion of Firms with Sales in More Than One Industry	Firms	Firm-year Obs.
11	0.74	2.02	0.42	0.57	76	797
21	0.58	1.38	0.21	0.34	1419	13296
22	0.59	1.99	0.47	0.66	466	6838
23	0.83	1.82	0.44	0.70	292	2823
31	0.88	1.69	0.34	0.59	703	7301
32	0.81	1.56	0.28	0.45	2622	27053
33	0.74	1.59	0.31	0.53	4289	47992
42	0.82	1.63	0.34	0.59	721	7081
44	0.71	1.35	0.22	0.45	488	5283
45	0.65	1.32	0.20	0.43	391	3419
48	0.58	1.61	0.31	0.52	484	5417
51	0.63	1.59	0.31	0.51	1680	16176
52	0.63	1.39	0.22	0.43	746	8838
53	0.76	1.42	0.25	0.50	1015	9377
54	0.75	1.66	0.35	0.60	483	4345
56	0.78	1.69	0.33	0.53	97	835
61	0.70	1.38	0.25	0.55	441	4142
62	0.78	1.51	0.31	0.56	173	1596
71	0.87	1.40	0.22	0.46	424	4181
72	0.78	1.46	0.29	0.48	85	738
81	0.98	1.50	0.17	0.37	550	4404
99	0.74	2.02	0.42	0.57	76	797
		Final Sample				
NAICS 2-Digit Code	Ratio of Sales in Other Inds. over Sales in Primary Ind.	Average Number of Sub-inds. for Firms-years	Firms-years with Sale in Other Inds. to Total Firms	Portion of Firms with Sales in More Than One Industry	Firms	Firm-year Obs.
11	0.87	1.59	0.29	0.29	7	17
21	0.78	2.02	0.41	0.50	31	98
22	0.96	1.89	0.50	0.54	6	28
23	0.84	2.26	0.53	0.62	14	66
31	0.78	1.73	0.38	0.52	107	629
32	0.79	1.53	0.30	0.45	551	3280
33	0.70	1.50	0.30	0.47	1589	9741
42	0.82	1.64	0.37	0.48	109	570
44	0.65	1.26	0.18	0.31	252	1900

33

		Final Sample				
NAICS 2-Digit Code	Ratio of Sales in Other Inds. over Sales in Primary Ind.	Average Number of Sub-inds. for Firms-years	Firms-years with Sale in Other Inds. to Total Firms	Portion of Firms with Sales in More Than One Industry	Firms	Firm-year Obs.
45	0.57	1.27	0.18	0.37	172	1125
48	0.89	1.22	0.22	0.56	5	9
51	0.48	1.33	0.22	0.40	99	425
52	0.64	1.32	0.25	0.40	94	607
53	0.83	1.55	0.33	0.52	88	413
54	0.84	1.52	0.35	0.45	53	220
56	0.74	2.16	0.55	0.65	9	31
61	0.70	1.48	0.33	0.56	59	267
62	0.77	1.32	0.19	0.23	31	189
71	0.87	1.26	0.18	0.37	196	1296
72	0.94	1.26	0.13	0.13	9	31
81	1.00	1.47	0.26	0.33	63	218
99	0.87	1.59	0.29	0.29	7	17

Table 2: (Continued)

Table 3: Variable definition and data sources.

Variable	Definition	Data Source
Tobin's q	((Annual calendar closing stock price * common shares outstanding) + total assets – total equity) / total assets	Compustat
ROA	Net income / total assets	Compustat
ROS	Net income / sales	Compustat
Total Profits	Net income	Compustat
OPEX to Sales	(Cost of goods sold + selling and general administrative expenses) / sales	Compustat
COGS to Sales	Cost of goods sold / sales	Compustat
Digital Proximity (dc)	Industry-level digital proximity is measured as: $$DC_i = \frac{1}{1 + \text{(average shortest path to IT industries)}}$$ We then form the firm-level digital proximity measure: $$dc_i = \frac{\sum_{k=1}^{n} DC_k \, Sales_k}{n}$$ where n is the total number of industries in which firm i reports sales.	Compustat Segments
Ind. Avg. Closeness Centrality	Average value of closeness centrality of firms reporting that industry as their primary industry within the industry space network. Closeness centrality of firms is measured as the weighted average closeness centrality of each industry in which firms report sales (Freeman 1979). Closeness centrality is the average shortest path from a node to all other nodes in the network.	Compustat Segments
R&D Intensity	R&D expenditure / sales	Compustat
Advertising Intensity	Advertising expenditure / sales	Compustat
Total Assets	Total assets, as reported.	Compustat
Intangible Assets	Total intangible assets, as reported.	Compustat
Diversification	Sum of squares (Firm's sales in industry i / Total sales in industry i) / log(Firm's sales in industry i / Total sales in industry i) for each industry i in which the firm operates (Jacquemin and Berry 1979).	Compustat Segments
HHI	Sum of square market share of each firm i in industry j (Hou and Robinson 2006).	Compustat
Weighted Market Share	(Firm's sales in industry i / Total sales in industry i) * (Firm's sales in industry i / Firm's total sales)	Compustat Segments

Table 4: Correlations and descriptive statistics for the final sample. N = 20,215									
		1	2	3	4	5	6	7	8	9
1	log(q)	1								
2	Digital Proximity	0.07	1							
3	ROA	-0.31	-0.10	1						
4	ROS	-0.32	-0.08	0.78	1					
5	ROI	0.02	-0.08	0.52	0.39	1				
6	ROE	0.08	-0.07	0.30	0.24	0.73	1			
7	Profitability	0.09	-0.06	0.28	0.21	0.28	0.23	1		
8	COGS/Sale	-0.15	-0.08	-0.21	-0.24	-0.12	-0.07	-0.13	1	
9	OPEX/Sale	0.31	0.08	-0.78	-0.91	-0.43	-0.28	-0.27	0.26	1
10	Ind. Avg. log(q)	0.16	0.13	-0.12	-0.13	-0.06	-0.03	0.03	-0.17	0.12
11	R&D	0.27	0.08	-0.44	-0.67	-0.26	-0.18	-0.11	0.06	0.62
12	Advertising	0.15	0.00	-0.26	-0.35	-0.16	-0.11	-0.04	-0.04	0.35
13	Ind. Avg. Closeness	0.04	-0.02	0.03	0.02	0.00	0.00	-0.04	-0.04	-0.01
14	Intangible Assets	-0.03	-0.01	0.17	0.14	0.13	0.11	0.63	-0.13	-0.22
15	Total Assets	-0.18	-0.06	0.48	0.37	0.26	0.17	0.61	-0.10	-0.47
16	Diversification	-0.13	-0.15	0.12	0.11	0.07	0.04	0.21	0.03	-0.14
17	HHI	-0.08	-0.17	0.05	0.06	0.05	0.03	0.00	0.15	-0.05
18	Market Share	-0.08	-0.16	0.20	0.17	0.16	0.13	0.47	0.07	-0.20
	Mean	0.61	0.24	-0.10	-0.26	-0.06	-0.03	49.40	0.58	1.06
	Std. Dev.	0.64	0.13	0.34	0.90	0.40	0.52	129.4	0.22	0.47
		10	11	12	13	14	15	16	17	18
10	Ind. Avg. log(q)	1								
11	R&D	0.14	1							
12	Advertising	-0.03	0.25	1						
13	Ind. Avg. Closeness	-0.09	0.01	0.06	1					
14	Intangible Assets	0.13	-0.07	-0.02	-0.07	1				
15	Total Assets	0.00	-0.20	-0.09	-0.08	0.62	1			
16	Diversification	0.06	-0.10	-0.07	-0.06	0.30	0.27	1		
17	HHI	-0.08	-0.10	-0.04	-0.11	0.00	-0.03	0.14	1	
18	Market Share	-0.11	-0.14	0.00	-0.08	0.40	0.51	0.23	0.32	1
	Mean	1.83	0.12	0.03	0.05	159.9	5.02	0.13	0.32	0.06
	Std. Dev.	1.19	0.36	0.04	0.01	371.4	2.36	0.26	0.24	0.13

Table 5: Fixed-effects (FE) and OLS model results for the effect of digital proximity on Tobin's q. Dependent is the logarithm of Tobin's q: log(q). Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Variable	(1) FE: log(q)	(2) FE: log(q)	(3) FE: log(q)	(4) OLS: log(q)	(5) OLS: log(q)	(6) OLS: log(q)
Digital Proximity	0.191***	0.183***		0.161**	0.552***	
	(0.0580)	(0.0292)		(0.0660)	(0.0334)	
log(Ind. Avg. Tobin's q)	0.0156***		0.0142***	0.0166***		0.0160***
	(0.00431)		(0.00434)	(0.00582)		(0.00580)
R&D Intensity	0.0528***		0.0519***	0.250***		0.249***
	(0.0196)		(0.0194)	(0.0183)		(0.0184)
Advertising Intensity	0.252		0.252	1.036***		1.056***
	(0.243)		(0.243)	(0.240)		(0.238)
Ind. Avg. Closeness Centrality	0.668		0.727	1.224		2.078
	(1.744)		(1.634)	(2.146)		(1.962)
Intangible Assets	-5.40e-05**		-5.18e-05*	8.78e-05***		9.06e-05***
	(2.75e-05)		(2.69e-05)	(2.76e-05)		(2.75e-05)
log(Total Assets)	-0.169***		-0.166***	-0.0561***		-0.0567***
	(0.0121)		(0.0119)	(0.00625)		(0.00619)
Diversification	-0.0434		-0.0771**	-0.189***		-0.215***
	(0.0320)		(0.0305)	(0.0335)		(0.0313)
HHI	-0.0590		-0.0704*	-0.143***		-0.157***
	(0.0390)		(0.0382)	(0.0395)		(0.0389)
Weighted Market Share	0.107		0.0714	0.503***		0.478***
	(0.0788)		(0.0756)	(0.0769)		(0.0731)
Constant	1.635***	0.350***	1.722***	0.792***	0.338***	0.959***
	(0.0983)	(0.0244)	(0.0923)	(0.0949)	(0.0299)	(0.0757)
Year Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes
Firm Fixed-Effects	Yes	Yes	Yes	No	No	No
Industry Fixed-Effects	No	No	No	Yes	Yes	Yes
Observations	20,215	126,630	20,771	20,215	139,258	25,170
Number of Unique Firms	3625	14286	3673	3625	15970	4516
R-squared	0.181	0.048	0.178	0.248	0.185	0.249
F stat	53.18***	180.9***	56.32***	45.07***	204.7***	62.84***

Table 6: Effect of digital proximity on alternative measures of performance and cost ratios, with fixed-effects (FE) panel regression models. The direction and statistical significance of the coefficient estimates for digital proximity in the models in this table are consistent with those obtained using Tobin's q as the dependent variable. Models in this table show the results of testing Model (1) in Table 5 using return on assets (ROA), return on sales (ROS), return on investments (ROI), return on equity (ROE), total profits, ratio of operational expenditures (OPEX) over sales, and ratio of cost of goods sold (COGS) over sales as dependent variables. The results presented in this table support our main results. The results in Models (1) to (5) show a positive and significant relationship between digital proximity and ROA, ROS, ROI, ROE, and total profits. In line with our overall findings, the results of Models (6) and (7) show a negative and significant relationship between digital proximity and the two cost ratios, COGS over sales and OPEX over sales. Dependent variables are defined in Table 3. Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Variable	FE: ROA	FE: ROS	FE: ROI	FE: ROE	FE: Prof.	FE: OPEX/Sale	FE: COGS/Sale
Digital Proximity	0.121***	0.206**	0.134***	0.131*	19.86*	-0.108***	-0.0568**
	(0.0304)	(0.0880)	(0.0508)	(0.0707)	(10.24)	(0.0367)	(0.0271)
Ind. Avg. Dep. Var.	0.000746	0.000225	0.00784	0.0222***	0.138***	-5.52e-05	-0.00128***
	(0.000812)	(0.000566)	(0.00485)	(0.00597)	(0.0211)	(0.00104)	(0.000465)
R&D Intensity	-0.107***	-1.060***	-0.123***	-0.103***	-2.928**	0.442***	0.214***
	(0.0152)	(0.0500)	(0.0218)	(0.0299)	(1.465)	(0.0297)	(0.0172)
Advertising Intensity	-1.341***	-4.784***	-1.858***	-2.237***	-127.2***	2.845***	0.143
	(0.151)	(0.436)	(0.219)	(0.291)	(24.97)	(0.220)	(0.135)
Ind. Avg. Closeness Centrality	-0.459	-0.563	-0.592	-0.134	-119.2	0.806	-0.377
	(0.934)	(1.300)	(1.112)	(1.962)	(289.2)	(0.750)	(0.513)
Intangible Assets	-9.67e-06	-4.50e-05***	-6.11e-06	-1.44e-05	0.0499***	-3.51e-06	-6.13e-06
	(7.23e-06)	(1.55e-05)	(1.00e-05)	(1.68e-05)	(0.00810)	(1.03e-05)	(6.69e-06)
log(Number of Employees)	0.0448***	0.0813***	-0.0118	-0.0318***	6.374***	-0.0610***	-0.00935**
	(0.00540)	(0.0129)	(0.00814)	(0.0110)	(1.736)	(0.00692)	(0.00444)
HHI	0.00388	0.0842*	-0.0416	-0.00970	-9.833	-0.0345	0.0330**
	(0.0220)	(0.0505)	(0.0340)	(0.0412)	(7.252)	(0.0241)	(0.0150)
Weighted Market Share	-0.0404	-0.0318	0.00771	-0.0398	97.28***	0.00723	0.0177
	(0.0258)	(0.0463)	(0.0404)	(0.0624)	(28.05)	(0.0288)	(0.0206)
Constant	-0.102**	-0.0913	-0.0349	0.0541	-16.13	0.927***	0.524***
	(0.0437)	(0.0773)	(0.0573)	(0.121)	(43.01)	(0.0505)	(0.0441)
Firm and Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	23,583	23,585	23,576	23,582	23,585	22,442	23,586
R-squared	0.074	0.364	0.037	0.023	0.097	0.309	0.140
Unique Firms	4,005	4,005	4,004	4,004	4,005	3,781	4,005
F stat	12.68***	28.89***	7.370***	6.097***	10.34***	18.59***	6.548***
df	36	36	36	37	36	36	36

Table 7: Results of testing the models on a sample of only diversified firms, Model (1), and only non-diversified firms, Model (2). These findings suggest that our main results are not driven by firms' overall levels of general diversification; rather, they are driven by firms' ability to diversify specifically toward more digital industries. Further, we tested two alternative models, one with the digital proximity measure obtained using the (inverse) mean path-length to digital industries (instead of the shortest path to digital industries) in Model (3), and the other with the ratio of firms' digital proximity over its industry average in Model (4). The results of Models (3) and (4) are in line with our main results. Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Variable	(1) FE: Diversified	(2) FE: Non-Diversified	(3) FE: Mean Weight	(4) FE: Relative to Ind.
Digital Proximity	0.256*	0.230***	0.689***	0.331***
	(0.151)	(0.0621)	(0.167)	(0.0861)
Controls	Yes	Yes	Yes	YEs
Constant	0.624***	1.516***	1.317***	1.463***
	(0.214)	(0.172)	(0.191)	(0.193)
Firm and Year FE	Yes	Yes	Yes	Yes
Observations	5,442	16,838	22,280	22,280
R-squared	0.176	0.155	0.152	0.152
F stat	5.340***	39.71***	52.11***	51.97***
Unique Firms	1088	3317	3794	3794
df	37	36	37	37

Table 8: First stage fixed-effects regression for the instrumental variable analysis. Robust standard errors in parentheses (*** $p<0.01$, ** $p<0.05$, * $p<0.1$).

Variable	(1) Digital Proximity	(2) Digital Proximity	(3) Digital Proximity	(4) Digital Proximity
Neigh. Distance to NAICS 541519 Computer-related Services	0.0141***	0.0178***		
	(0.000584)	(0.000590)		
Neigh. Distance to NAICS 511210 Software Publishers	-0.0148***	-0.0141***		
	(0.000542)	(0.000540)		
Neigh. Distance to NAICS 541512 Computer Systems Design Services	-0.00367***	-0.00340***		
	(0.000536)	(0.000552)		
Past Yr. Neigh. Distance to NAICS 541519 (Computer-related Services)	-0.0106***		0.00407***	
	(0.000570)		(0.000621)	
Past Yr. Neigh. Distance to NAICS 511210 (Software Publishers)	0.00659***		-0.00738***	
	(0.000527)		(0.000568)	
Past Yr. Neigh. Distance to NAICS 541519 (Computer Systems Design Svcs)	0.00954***		0.00762***	
	(0.000532)		(0.000584)	
Past Yr. Digital Proximity	0.539***			0.539***
	(0.00296)			(0.00297)
2nd Past Yr. Digital Proximity	0.0652***			0.0659***
	(0.00296)			(0.00297)
Constant	0.0713***	0.215***	0.202***	0.0752***
	(0.00144)	(0.00147)	(0.00151)	(0.00105)
Observations	122,296	158,406	139,607	122,717
R-squared	0.381	0.051	0.052	0.373
Number of Unique Firms	14,160	16,312	15,406	14,180
F stat	1956***	244.6***	228***	2303***
Number of Unique Firms	14160	16312	15406	14180
df	14193	16342	15435	14207

38

Table 9: Instrumental variables analysis using Panel 2SLS estimator. Model (1) uses the proximity of an industry's neighbors (second-order proximity) to three IT industries with NAICS codes of 511210 (software industries), 541512 (computer system design services), and 541519 (other computer services) as instruments. Model (2) use the same instruments from Model (1) in addition to their first logs. Model (3) uses the first and second logs of digital proximity as instruments. All models show positive and significant results. Moreover, the magnitude of the digital proximity coefficient in these models is higher than in the main model reported in Table 5, suggesting that if there are biases introduced by unobserved factors, they are in the opposite direction of our hypothesized relationship. Over-identification tests (insignificant Hansen J statistics), week identification tests (Cragg-Donald Wald statistics and the Stock-Yogo critical values), and the under-identification test (Anderson canonical correlation LM statistic) do not suggest that the instrumental variables are invalid. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

	(1)	(2)	(3)
Variable	log(q)	log(q)	log(q)
Digital Proximity	0.543***	0.623***	0.211***
	(0.180)	(0.180)	(0.0711)
log(Ind. Avg. Tobin's q)	0.0166***	0.0166***	0.0129***
	(0.00389)	(0.00396)	(0.00393)
R&D Intensity	0.0525***	0.0285**	0.0319**
	(0.0115)	(0.0129)	(0.0137)
Advertising Intensity	0.255*	0.225	0.0372
	(0.140)	(0.147)	(0.155)
Ind. Avg. Closeness Centrality	0.472	-0.372	-1.243
	(1.329)	(1.402)	(1.420)
Intangible Assets	-5.92e-05***	-7.20e-05***	-7.37e-05***
	(1.76e-05)	(1.82e-05)	(1.84e-05)
log(Total Assets)	-0.172***	-0.169***	-0.171***
	(0.00562)	(0.00598)	(0.00620)
Diversification	-0.0485**	-0.0403*	-0.0564***
	(0.0221)	(0.0226)	(0.0212)
HHI	-0.0386	-0.0267	-0.0433*
	(0.0249)	(0.0258)	(0.0249)
Weighted Market Share	0.119**	0.114**	0.0665
	(0.0515)	(0.0540)	(0.0539)
Firm and Year Fixed-Effects	Yes	Yes	Yes
Observations	21,612	19,997	18,226
R-squared	0.147	0.144	0.154
Number of Unique Firms	3,175	2,956	2,693
F stat	85.67	80.48	78.24
Number of Unique Firms	3175	2956	2693
df	3213	2993	2729
Hansen p-value	0.871	0.178	0.805
Hansen J stat	0.275	7.625	0.0612
Cragg-Donald Wald F statistic	289.138***	146.249***	3811.531***
Stock-Yogo 5% maximal IV relative bias	13.91	19.28	NA
Stock-Yogo 10% maximal IV size	22.30	29.18	19.93
Anderson canon. corr. LM statistic (p-value)	830.157 (0.000)	836.482 (0.000)	5121.705 (0.000)

Table 10: Results of the estimated logit model. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.	
	(1)
Variable	Digital Proximity (binary form)
Software Alliances	0.229***
	(0.0623)
Software Development Services	0.00910
	(0.0911)
Total Alliances	-0.102***
	(0.0215)
Software Partners	-0.107
	(0.0720)
Marketing Services	-0.0477
	(0.0338)
Joint Ventures	0.111***
	(0.0292)
Software Targets	-0.268***
	(0.0751)
Total Targets	-0.0541***
	(0.00924)
Ind. Avg. Software Targets	0.00388***
	(0.000274)
Constant	-0.000231
	(0.0311)
Observations	7,397
Chi-squared	436.21
Pr > Chi-squared	0.000
Pseudo R-squared	0.02

Table 11: Results of the propensity score matching approach with different techniques and specifications. Column (1) reports the result of Kernel matching. Nearest neighbor matching results are reported with 2 neighbors and a caliper of 0.005 in Column (2), 10 neighbors and a caliper of 0.01 in Column (3), and 10 neighbors and a caliper of 0.005 in Column (4).				
	(1)	(2)	(3)	(4)
Matching Specifications	Kernel	NN-Matching	NN-Matching	NN-Matching
Mean Outcome - Treated	0.4	0.4	0.4	0.4
Mean Outcome - Control	0.17	0.12	0.15	0.15
Difference	0.23	0.28	0.25	0.25
S.E.	0.01	0.05	0.02	0.02
t-statistic	15.92	6.05	11.15	11.13
Mean Percentage Bias	1.17	0.9	0.78	0.71
Median Percentage Bias	0.87	0.57	0.64	0.76
All Biases below 5%	YES	YES	YES	YES
N	7394	7394	7396	7394
N Treatment	3905	3905	3907	3905
N Control	3489	3489	3489	3489
Off Support	3	3	1	3
Gamma	1.8	4.2	1.8	1.8

Table 12: Descriptive statistics before and after matching. Note that there is no difference between the mean values for the treatment group before and after matching. The reported t-statistic and p-value are the results of a t-test between the mean treatment and mean control after matching values.

	Mean Treatment	Mean Control Before Matching	Mean Control After Matching	t-statistic	p-value
Software Alliances	0.079	0.161	0.076	0.21	0.832
Software Development Services	0.023	0.059	0.027	-0.46	0.646
Total Alliances	0.475	1.22	0.567	-1.6	0.109
Software Partners	0.064	0.157	0.067	-0.13	0.889
Marketing Services	0.104	0.313	0.129	-1.36	0.173
Joint Ventures	0.175	0.337	0.19	-0.53	0.598
Software Targets	0.054	0.139	0.051	-0.53	0.599
Total Targets	0.98	1.908	0.979	-0.12	0.901
Ind. Avg. Software Targets	85.055	47.751	88.199	-1.2	0.229

Table 13: Results of the shuffling test. Column (1) reports the mean and standard error (S.E.) of the digital proximity coefficient when randomly shuffling digital proximity scores and estimating the coefficient 100 times. The statistically insignificant results (comparing the mean to the S.E.) presented in this table show that our identified relationship is not attributed to chance or common industry trends. Column (2) reports the same when shuffling within industries. Finally, Column (3) shows the results of randomly assigning digital proximity scores within each year and industry (i.e., preserving year and industry variables). Pseudo p-values are constructed based on the percentage of times that a z-test reported statistically insignificant differences between the estimated model and the original estimation in Table 5. Highly significant pseudo p-values for all models, therefore, show that the coefficient estimates for digital proximity obtained through shuffling were significantly different from those reported in Table 5. Therefore, the main results are not attributed to chance and common industry trends.

	(1)	(2)	(3)
Mean	.0013	.0006	.0012
S.E.	.0015	.0029	.0031
Pseudo p-value	0.00	0.00	0.00

About the Authors

Pouya Rahmati is an Assistant Professor of Management Information Systems at the Terry College of Business, University of Georgia. His research, which focuses on co-developing digital products in complex economic and organizational environments, employs empirical analyses to identify and quantify the drivers of firms' success in developing digital products and services. He has presented his works in many prestigious conferences in the field of Management Information Systems. He completed his PhD at the University of Illinois at Chicago; and prior to entering academia, he worked as a software engineer, product manager, and IT-business analyst.

Ali Tafti is an Associate Professor of Information & Decision Sciences in the College of Business Administration at the University of Illinois at Chicago. His research interests include economic and strategic impacts of information technology investment, social and collaborative networks, and causal inference methods. Dr. Tafti enjoys teaching graduate-level courses on econometrics, causal inference, and social network analysis. His work has appeared in journals such as *Management Science*, *MIS Quarterly*, *Information Systems Research*, *MIT Sloan Management Review*, and *PLoS One*. He has previously served on the faculty at the University of Illinois at Urbana–Champaign. He completed his PhD at the Ross School of Business at the University of Michigan; and he previously worked as a software engineer and analyst in the digital business consulting and financial industries.

J. Christopher Westland is Professor in the Department of Information & Decision Sciences at the University of Illinois. He holds a BA in Statistics and an MBA in Accounting and received his PhD in Computers and Information Systems from the University of Michigan. His most recent book Audit Analytics provides a road-map and software for the application of statistical and machine learning in the automation of public accounting. He is the author of numerous academic papers and books, including Financial Dynamics (Wiley 2003); Valuing Technology (Wiley 2002); Global Electronic Commerce (MIT Press 2000); Global Innovation Management (Palgrave Macmillan 2008; Springer 2017); Red Wired: China's Internet Revolution (Marshall Cavendish, 2010); Structural Equation Modeling (Springer 2015, 2nd edition 2019) and Audit Analytics (Springer 2020). He has served on the editorial boards of numerous scholarly journals and is current Editor-in- Chief of the premier Electronic Commerce Research journal. He has served on the faculties at the University of Michigan, University of Southern California, Hong Kong University of Science and Technology, Tsinghua University, Nanyang Technological University, and other academic institutions. He has been designated by China as a High-level Foreign Expert under the 1000-Talents Plan. He has advised on valuation and technology strategy for Microsoft, Intel, Motorola, V-Tech, Aerospace Corporation, IBM, Pacific Bell, and other technology firms.

César A. Hidalgo is a Chilean-Spanish-American scholar known for his contributions in economic complexity, data visualization, and artificial intelligence. He holds a PhD in Physics from the University of Notre Dame and a Bachelor in Physics from Universidad Católica de Chile. Dr. Hidalgo currently holds a Chair at the [Artificial and Natural Intelligence Toulouse Institute](#) (ANITI), where he leads the Collective Learning group. He is also an Honorary Professor at the University of Manchester and a Visiting Professor at Harvard's School of Engineering and Applied Sciences. Hidalgo if the former head of [MIT](#)'s Collective Learning group. He was part of MIT's faculty from 2010-2019 (first Assistant and then Associate Professor). Prior to joining MIT, Hidalgo was a research fellow at Harvard. Hidalgo is also a founder of [Datawheel](#), an award winning company specialized in the creation of data distribution and visualization systems.

Online Appendix: When All Products Are Digital: Complexity and Intangible Value in the Ecosystem of Digitizing Firms

Pouya Rahmati (University of Georgia; rahmati@uga.edu)
Ali Tafti (University of Illinois at Chicago; atafti@uic.edu)
Christopher Westland (University of Illinois at Chicago; westland@uic.edu)
Cesar Hidalgo (University of Toulouse; cesar_hidalgo@univ-toulouse.fr)

Appendix A: Methodology Note

Background of the Empirical Method

We build our measurement of the digital proximity construct on some key insights in Porter (2008) regarding the importance of a firm's positioning in its environment for a competitive advantage and in Porter and Siggelkow (2008) regarding the dynamic and complex nature of this environment. We conceptualize the industry space environment and extend Porter's model to an environment made complex by a rich network of contextual complementary interactions. This allows our measurement to address how firms identify and occupy profitable positions when patterns of digitization are constantly changing the competitive landscape and the pay-offs associated with creating different sets of complementarities (Lee et al. 2010; Tanriverdi et al. 2010).

We develop the competitive landscape based on the theory of economic complements (Milgrom and Roberts 1990). Based on this theory, firms are rational actors that only diversify into a limited set of industries, which allows them to enhance the value they generate from their existing resource and capability structures. Therefore, if a firm has diversified into a certain set of different industries, it can be inferred that the resources and capabilities required for operating in those industries are economic complements. Accordingly, a firm reporting sales in two industries can signal potential complementarities between those two industries, and if many firms do

43

the same, this can be considered a sign of established complementarities between industries.

The industry space is based on a methodology widely used when analyzing the economic complexity of nations. Hidalgo et al. (2007) use countries' exports to develop a product space network over which they measure the complexity of different products based on their positioning in the network. To remove marginal exports from their analysis, they consider only the exports in which the exporting country has a revealed comparative advantage (RCA). The RCA is measured using the ratio of the share of the export of a given product from a given country's exports over the share of the export of that product from the global economy. Only exports with an RCA greater than one are considered in their analysis.

Based on RCA measures, Hidalgo et al. (2007) form a product space network in which two products are linked if at least one country has an RCA greater than one for both products. They define $\phi_{i,j}$ as the proximity between products i and j and measure it as:

$$\phi_{i,j} = \min[P(RCA_i|RCA_j), P(RCA_j|RCA_i)] \qquad \text{Eq. (A1)}$$

They develop the economic complexity measure in relation to the proximity of countries' exports to the core of the product space network—the core being where more sophisticated products are located. Further, they use the product space network to analyze the evolution of countries' productive structures. They do not develop periodical time-varying product spaces; rather, they use only one product space for their analysis. They look at the set of products for which countries have an RCA and how their export structures evolve into the products that are in close proximity to those products. Their analysis shows that the structure of the product space limits

many countries' ability to approach the core, and that the exports of more developed economies are located around the core of this network.

Industry Space

We develop the industry space using all groups of 6-digit NAICS industry classifications reported in the Compustat Segments database. We use industry classifications because this level of detail represents the products and services firms offer. In this network, each industry is a node, and the weight of the link between two industries i and j, $\phi_{i,j}$, is the minimum of two conditional probabilities: the probability that a firm reporting sales in industry i also reports sales in industry j, and vice versa.

$$\phi_{i,j} = \min[P(x_i|x_j), P(x_j|x_i)] \qquad \text{Eq. (A2)}$$

To capture the dynamics of the industry space, we develop different versions of the industry space for different time periods. To facilitate analysis, we develop yearly industry space networks using the sales reported in each year. To create visualizations, we develop periodic industry space networks for five five-year periods (i.e., periods one to five) between 1990 and 2014.

The industry space is a sparse network that actually became sparser until 2010 (see Figure A1b), as evident in the increasing ratio of nodes to edges. In period five (2010-2014), there are 7,142 links between 802 industries. Meanwhile, the weight of the links between industries has been growing, as well (see Figure A1a). The percentage of links with a weight higher than 0.2 has increased from around 10% in period one to around 20% in period five. In the industry space for period five, around 52% of the nodes have a weight lower than 0.022, 50% have a weight lower than 0.060, and around 75% have a weight lower than 0.166.

The industry space in Figure 1 is developed with undirected links and weights that are based on the number of firms reporting sales. To examine the robustness of this procedure for weighting network links, we develop five alternative weighting schemes for network links. First, we construct an alternative industry space in which links are weighted by the proportional amount of sales in respective industries. Next, for both the original and alternative industry spaces, we construct versions based on directed links. The directed version of the weighting scheme yields two distinct link weights for each pair of industries. Thus, for the original and alternative industry spaces, we have four alternative *directed* digital proximity measures in addition to two undirected digital proximity measures.

Digital Proximity

We first measure the digital proximity index in each industry by calculating the shortest-path network distance from each industry to software-producing industries. For a definition of digital (i.e. IT-producing) industries, see Table A1.To calculate shortest paths using calculated proximities in the industry space, we transform the proximies (edge weights) to $\log\left(\frac{1}{\phi_{i,j}}\right)$ to assign a lower cost to paths with higher proximity in the industry space; then, we use Dijkstra's algorithm to compute the shortest paths between nodes. We then compute the maximum shortest path length to IT industries. The digital proximity of industry *i* is:

$$DC_i = \frac{1}{1+ \text{(maximum shortest path to IT industries)}} \qquad \text{Eq. (A3)}$$

We also consider the three alternative industry space constructions: the alternative undirected industry space and the two alternative directed versions of industry space—each of which yield two distinct directed shortest path lengths for each pair of industries.

Using the industry-level measure of digital proximity, we then extend this measure to the firm level (dc_i) by calculating the average digital proximity of the n industries that a firm operates in, weighted by the amount of sales that the firm reports in each of those industries ($Sales_k$).

$$dc_i = \frac{\sum_{k=1}^{n} DC_k Sales_k}{n} \qquad \text{Eq. (A4)}$$

Using the digital proximity index allows a comprehensive look at firms' digitization processes, as well as their general diversification. Figure A2 shows how firms that report the same industry as their primary industry might choose different paths in the industry space. Firms in both manufacturing and credit intermediation industries show a diversification trend toward the bottom right of the industry space, where IT industries are located. Firms that drive this shift create direct complementarities with IT industries and have higher digital proximity indices.

Propensity Score Matching with Alternative Covariates

In order to ensure the robustness of our results to the covariates selected, we repeat the propensity score matching procedure for an alternative set of covariates. These covariates include the sales efficiency, R&D intensity, intangible assets, and net margin associated with firms' levels of digitization, all of which we have access to from the Compustat dataset (Chang and Gurbaxani 2012; Im et al. 2001; Kleis et al. 2012); we also use the industry-level average of digital proximity. Taking advantage of the panel nature of our data, we construct our covariates as average values between the years 2013 and 2015, construct our treatment variable at year 2015, and construct our outcome variable at year 2016. We report the result of the logit regression model in Table A2. Table A3 reports the results of our propensity score matching analysis using the nearest neighbor matching technique, with a caliper of 0.01 in Column (1) and 0.001 in Column (2). Both results are consistent with the results generated when

using the covariates developed based on strategic alliances, mergers, and acquisitions.

We report the relevant mean and median biases and critical Gamma values in Table

A3 and the descriptive statistics for our covariates before and after matching in Table

A4, further establishing the robustness of the alternative propensity score matching

results.

Figure A1: Cumulative weight distribution (a) and changes of the node/edge ratio (b) in the industry space.

| (a) | (b) |
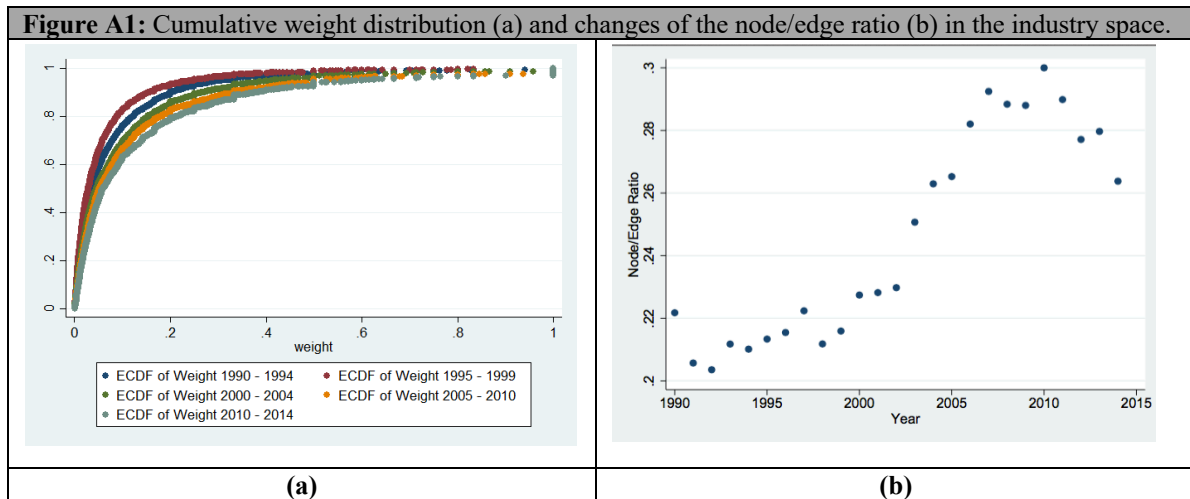
Figure A2: Industry space diversification of firms from 2010 to 2014. Machinery manufacturing industries (NAICS 333) are reported in the first row; credit intermediation and related activities industries (NAICS 522) are reported in the second row. The red nodes in the networks are different industries that firms from each of these two industries have reported sales in. The IT sectors are generally located in the bottom right of the network diagrams.

| 2010 | 2011 | 2012 | 2013 | 2014 |
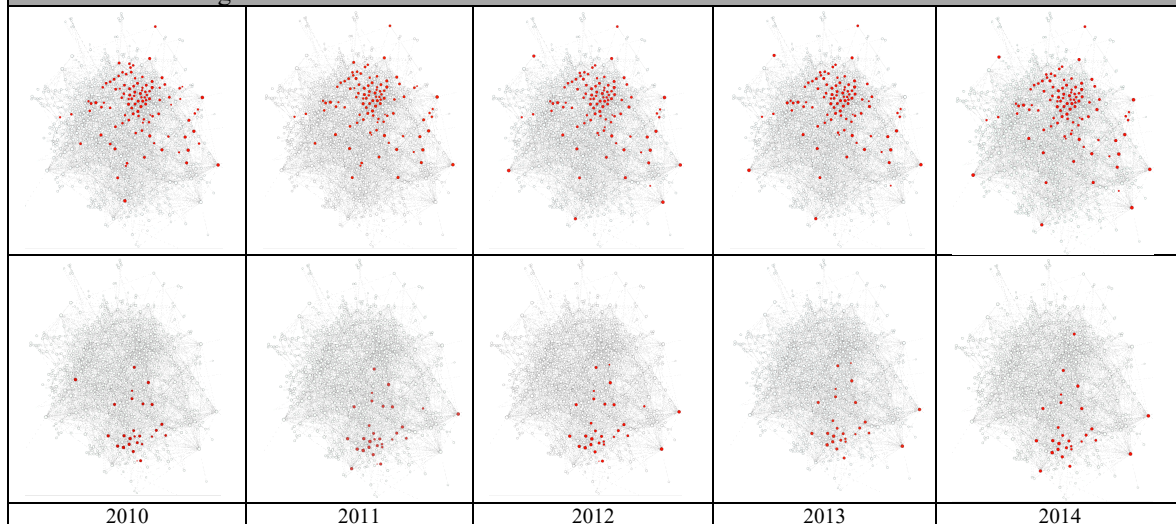
Table A1: IT-producing industries, excluded from the sample.		
NAICS Code	Description	Examples
5112	Software publishers	SAP, Oracle, Microsoft
5182	Data processing	ADP , Fiserv
5191	Information services	Google, Baidu, Alibaba, Facebook
5415	Systems design	IBM, Fujitsu, Accenture

48

Table A2: Results of estimated logit model. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.	
Variable	Digital Proximity (binary form)
Sales Efficiency	-0.0254
	(0.0699)
R&D Intensity	-0.463***
	(0.116)
Net Margin	-0.224***
	(0.0643)
Intangible Assets	0.000230**
	(9.53e-05)
Ind. Avg. Digital Proximity	-0.224***
	(0.0571)
Constant	-0.561***
	(0.154)
Observations	2,614
Chi-squared	56.13
Prob > Chi-squared	0.000
Pseudo R-squared	0.019

Table A3: Robustness test for propensity score matching. Column (1) includes the result of performing propensity score matching using all variables in Table A2, with nearest neighbor matching and a caliper of 0.01. Column (2) repeats the same procedure in Column (1) with a caliper of 0.001.		
	(1)	(2)
	NN-Matching	NN-Matching
Mean Outcome Treated	.22	.17
Mean Outcome Control	.14	.11
Difference	.07	.06
S.E.	.02	.02
t-statistic	3.13	2.58
Mean Percentage Bias	1.67	1.67
Median Percentage Bias	1.19	1.44
All Biases below 5%	YES	YES
N	2614	2428
N Treatment	646	551
N Control	646	551
Off Support	1968	1877
Gamma	1.5	1.4

Table A4: Summary statistics of the covariates before and after matching.					
	Mean Treated	Mean Control - Before Matching	Mean Control - After Matching	t-statistic	p-value
Sales Efficiency	0.785	0.772	0.785	0.2	0.79
R&D Intensity	9,285	0.474	0.274	0.18	0.856
Net Margin	-0.552	-0.738	-0.5001	-0.19	0.849
Intangible Assets	341.32	270.46	328.87	0.6	0.551
Ind. Avg. Dig. Prox.	2.436	2.69	2.417	0.3	0.762

Appendix B: Digital Proximity, Similarity to Digital Knowledge, Knowledge Complexity, and Digital Innovations

We assess the face validity of the digital proximity measure by comparing it to three measures of (1) similarity to digital knowledge, (2) knowledge complexity, and (3) digital innovations. First, we use the knowledge similarity measure developed by Farjoun (1994; 1998) (Figure B2a). Following this work, we use Occupational Employment Survey data from the Bureau of Labor Statistics (BLS) and measure the similarity of the knowledge structure between every 4-digit NAICS code and digital industries by calculating the Euclidian distances between them. We find that the knowledge similarity indices are highly correlated with industry-level digital proximity indices, suggesting the digital proximity measure captures the similarity between industry-level knowledge structures.

Second, we use the BLS dataset to form a directed bipartite network in which each industry code has directed edges pointing to the set of occupations that it reports. In order to construct the measure of knowledge complexity, we use the BLS Occupational Employment Statistics and develop a bipartite network in which nodes represent both the industries and the occupation classifications. In this directed network, there is a link between an industry and an occupation if the occupation is utilized by that industry. We measure link weights as the product of the percentage of the employees working in that occupation in that industry and their median wages in that industry. This helps normalize the effect of less knowledge-intensive occupations (typically earning lower than the median) that form large proportions of an industry's workforce, as well as more knowledge-intensive occupations (typically earning higher than the median) that form small proportions of an industry's workforce.

50

We use the Hyperlink-induced Topic Search (HITS) algorithm in order to assign a hub score to each industry (Kleinberg 1999). This algorithm was originally aimed at assessing the importance of the information that different webpages point to, based on the iterative assessment of the importance of other webpages pointing to the same information. This algorithm assigns two scores of hub and authority to each node in a directed network. The authority scores, representing the significance of the information, are set to be the sum of the hub scores of the nodes that point to them. The hub scores are set to the sum of the authority scores of the nodes that they point to. Therefore, a node that has a high hub score points to a set of nodes with high authorities, and a node that has a high authority score is pointed to by nodes considered information hubs. The algorithm starts by assigning an initial value of 1 for both hub and authority scores for all nodes, updates hub and authority scores iteratively, and normalizes the values by dividing each score by the square root of the sum of the squares of that same score after each iteration.

We use this algorithm by considering the authority scores of different occupation groups as a representation of the complexity level associated with them because high authority scores indicate the significant role of those occupation groups in a diverse set of industries. Similarly, we use industry hub scores as a measure of their knowledge structure complexity because high hub scores indicate the use of diverse and complex knowledge structures in an industry. By considering the actual importance and diversity of each occupation group, our approach allows us to go beyond more traditional measures of complexity and diversity, such as count-based measures, and provide a more accurate measure of the knowledge complexity of different industries. Figure B1 compares the two ego networks (of size two) for the apparel knitting mills industries (left) and computer systems design industries (right).

In these networks, the focal industries are indicated by the color red, the occupation groups by the color blue, and other industries by the color green. Industry node size is proportional to their hub scores, and the size of the occupation group nodes is proportional to their authority scores. Comparing the two networks reveals how a more complex industry gains higher hub scores through connections with occupations that have high authority scores, and how a more complex occupation group gains higher authority scores by being connected to a higher number of industries with high hub scores. The visualizations in Figure B1 are developed using hub and authority scores measured based on 2-digit OCC codes; in constructing the original measures, the most granular OCC codes are used.

Substantively, our knowledge complexity measure is similar to the diversity measure developed by Hidalgo and Hausmann (2009). The latter captures the diversity of countries' exports by looking at the set of products that they produce, and the other countries that produce those products, and so on. The HITS algorithm which produces our hub-score based measure of knowledge complexity has a similarly recursive structure, which is advantageous because it captures the flow of knowledge through indirect connections across the entire network. The knowledge complexity measure is highly correlated with the digital proximity index, confirming the complex nature of digital know-how (Figure B2b).

Third, we use the patents dataset developed by Kogan et al. (2017) to examine the relationship between the digital proximity measure and innovation and digital innovation more specifically. We use the patents data between years 1990 and 2010 and examine the association between digital proximity and firms' total number of patents, digital patents, and AI-related patents. We consider patent classes between 700 and 726 as digital patents, which include data processing, software, and computer

design patents. We consider patent class 706, "digital processing: artificial intelligence," as AI patents. The results of this analysis, reported in Table B1, show a significant and positive association between digital proximity and firms' general innovative output, as well as their digital innovations more specifically.

Figure B1: Ego (size 2) network of (a) apparel knitting mills industries and (b) computer systems design industries. In these network diagrams, red represents the focal industry, blue represents occupation groups, and green represent other industries. Industry size is proportional to their hub scores, and occupation group size is proportional to their authority scores. Comparing these two networks reveals how a more complex industry (e.g., computer systems design) gains higher hub scores through connections to a large set of occupation groups with higher authority scores.

(a)   (b)
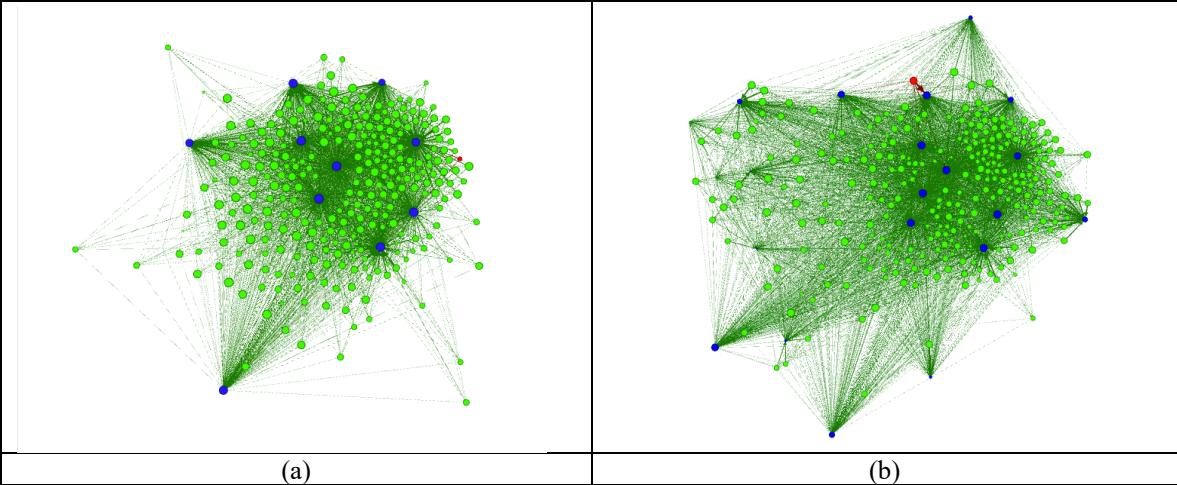
Figure B2: Relationship of digital proximity to a) knowledge similarity and to b) knowledge complexity. Figure B2a shows a positive correlation (a correlation of 0.621, significant at a 0.000 level) between (4-digit NAICS) industries' digital proximity indices and their levels of knowledge similarity with digital industries. Knowledge similarity scores are developed using the BLS occupational employment survey. This value is calculated as the average Euclidian distances between an industry and the four digital (IT-producing) industries, using a vector of major occupational groups with the percentage of employees working in each group as the weight for that group (the direction of the measure was inverted to assign higher similarity scores to lower distances). Figure B2b shows the positive correlation between digital proximity indices and knowledge complexity scores (a correlation of 0.382 at a 0.000 level). Knowledge complexity scores are developed by calculating the hub scores for each industry in a bipartite network, with each industry pointing to the different occupation groups in it. All measures are winsorized at 1st and 99th quantiles.

(a)   (b)

Table B1: Relationship between digital proximity and firms' innovation output in non-software firms. The dependent variable in 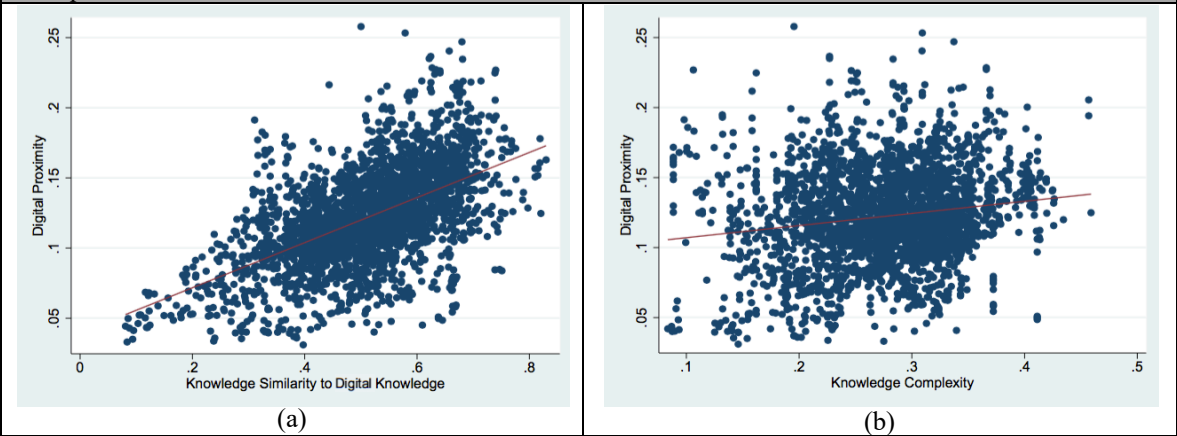Model (1) is log(number of patents). Models (2) and (3) examine the relationship between digital proximity and firms' ability to produce digital innovations. We consider patents in the 700 to 726 classes as digital patents. These classes include data processing, software, and computer systems design patents. In Model (2) the dependent variable is the total number of digital patents, while in Model (3) it is the ratio of digital patents to total patents. Models (4) and (5) look at the relationship between digital proximity and AI patents (patent class 706). The dependent variable in both models is the total number of AI patents. In Model (5) the sample is limited to firms with at least one digital patent. In all models, industry patents and industry software patents are the total number of patents and the total number of software patents in firms' 3-digit NAICS code industry. All dependent variables are the next year (first lead) values. The results remained consistent when using the same year values and next leads. Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Variable	(1) Patents	(2) Digital Patents	(3) Digital/ Total Patents	(4) AI Patents	(5) AI Patents
Digital Proximity	0.391***	14.23***	0.161***	0.239***	0.213**
	(0.138)	(5.365)	(0.0313)	(0.0739)	(0.103)
Industry Patents	6.88e-05***	9.53e-05	-4.33e-06***	-4.54e-06*	-2.35e-05*
	(8.76e-06)	(0.000128)	(1.35e-06)	(2.44e-06)	(1.31e-05)
Industry Digital Patents	-0.000138***	0.00236***	2.89e-05***	3.09e-05**	0.000126*
	(3.96e-05)	(0.000733)	(6.36e-06)	(1.29e-05)	(6.59e-05)
R&D Intensity	0.125***	1.256***	-0.00308	0.0256***	0.0794***
	(0.0189)	(0.339)	(0.00190)	(0.00678)	(0.0291)
log(Total Assets)	0.384***	1.843***	0.00291*	0.0265***	0.0712***
	(0.0149)	(0.298)	(0.00166)	(0.00569)	(0.0154)
Diversification	0.0943	5.536*	0.000543	0.0832**	0.224*
	(0.0648)	(2.912)	(0.00675)	(0.0345)	(0.115)
HHI	-0.0446	-0.376	-0.00301	-0.0516***	-0.153**
	(0.0819)	(1.076)	(0.00907)	(0.0185)	(0.0705)
Weighted Market Share	0.436***	-0.737	-0.0173	0.0889	0.183
	(0.166)	(3.425)	(0.0141)	(0.0579)	(0.194)
Constant	-1.051***	-11.94***	0.0385***	-0.166***	-0.391***
	(0.106)	(3.425)	(0.0142)	(0.0451)	(0.129)
Observations	14,108	14,108	14,108	14,108	3,567
R-squared	0.141	0.0651	0.0122	0.00269	0.0152
Wald chi2	1113	77.90	183.9	49.85	63.31
Number of Unique Firms	2308	2308	2308	2308	765
df	27	27	27	27	27
Year Dummies	Yes	Yes	Yes	Yes	Yes

Appendix C: Supplementary Robustness Tests

We examine the models for potential issues of multicollinearity, unobserved heterogeneity, and reverse causality.

First, we address the potential issue of multicollinearities. We start by testing our main models, in Table 5, with and without the control variables and the main independent variable. The coefficient estimates for the digital proximity variable and

54

the control variables remain consistent across different models, providing evidence for no significant multicollinearities in our main models. We further examine our models using alternative digital proximity measures and find similarity between the obtained results. Table C1 compares the results for the main digital proximity measure in Model (1) and the five alternative measures of digital proximity shown in Models (2) to (6). The results show similar magnitude of effects, with the two directed measures developed based on the shortest paths *from* IT slightly larger in magnitude and the results developed based on the amount of sales more significant (at $p < 0.01$ level). Finally, we run a variance inflation factor (VIF) diagnostic for our main model. These results show a mean VIF of 3.23, which is below a threshold of 10, and all of the VIF values below 10. This undermines potential concerns for multicollinearities.

Second, we explore the potential issue of unobserved heterogeneities. We provide additional tests—in addition to the instrumental variable technique, propensity score matching, shuffling techniques, and testing the models in different sub-samples based on firms' levels of diversification, as discussed earlier—to further investigate this issue. We start our investigation by comparing the OLS and fixed-effects results in order to address potential concerns regarding unobserved variables. Many unobserved factors such as firms' innovative culture or managerial practices that might affect both intangible value and digital proximity are relatively stable factors with little change over time. While these firm-specific factors are controlled by the use of firm fixed-effect models, comparing the results obtained from these models with those from OLS models could reveal the sensitivity of the identified relationship to these factors. Table 5 compares the OLS and fixed-effects results and shows that the magnitude and significance of the effect of digital proximity on

Tobin's q are similar, albeit somewhat higher in the OLS models. The similarities between OLS and fixed-effects model results suggest that unobserved firm-specific factors do not have a significant confounding influence on the effects of digital proximity on intangible value. We also compare other alternative model specifications with our fixed-effects results. Although the significant Hausman test statistic suggests that fixed-effects rather than random-effects models should be used for the hypothesis tests, the range of estimates generated by these two models can help demonstrate the robustness of our main coefficient estimates. While fixed-effects panel models utilized only the longitudinal variation of measures for each firm over time, random-effects models are efficient in the sense that they make use of the full range of between-firm and within-firm variation. Table C2 shows random-effects panel model results, as well as two panel generalized-estimating equation (GEE) with Huber-White robust standard errors. Here, Model (3) allows for exchangeable, and Model (4) allows for autoregressive within-group correlation structures. Coefficient estimates of Models (2), (3), and (4) in Table C2 are broadly similar to the consistent fixed-effects estimator in Model (1) in support of the hypothesis, despite the different underlying econometric assumptions distinguishing the models.

Finally, we further test our models for the potential influence of reverse causality, which may be present to the extent that firms' pre-existing intangible value allows them to enter new markets and increase their digital proximity. Although the instrumental variables and propensity score matching techniques, to some extent, alleviate the reverse causality concerns, we further examine the potential influence of this factor. We start by performing a formal Granger causality Wald test with six lagged values of Tobin's q and digital proximity. The test concluded that digital proximity does not Granger-cause Tobin's q (with a chi-square of 9.53 and p-value of

0.146). We further test several models to consider the effects of digital proximity and Tobin's q on lagged and lead values of each other. Table C3 shows results examining the effect of digital proximity on past, present, and future values of Tobin's q. The results in this table show different signs and magnitude of the effect for past and future values of Tobin's q. We find no significant relationship between digital proximity and past values of intangible value; however, the effect of digital proximity on future values of Tobin's q is largely positive and significant. We further address the reverse-causality concern by examining how firms' intangible value might affect their digital proximity levels. Model (1) in Table C4 presents the results of this analysis, showing no significant effect of present and past values of Tobin's q on future values of digital proximity. Granger reverse-causality can be demonstrated to be absent when $f(x_t|x_{t-1}, y_{t-1})$ equals $f(x_t|x_{t-1})$ where x is the set of covariates and y is Tobin's q; specifically, when y_{t-1} has no significant effect on the independent variable x_t when the past values of the covariates (e.g. x_{t-1}) are included (Greene 2003). Table C4 shows no significant effect on digital proximity for up to six lagged values of Tobin's q, and the results more generally suggest that lagged values of Tobin's q have no explanatory effect on digital proximity beyond the factors that our models already control for. Models (2) to (7) in this table show the effect of present and previous years of Tobin's q on present and future values of digital proximity. We find no significant effect of Tobin's q on digital proximity. The results reported in Table C4 support those reported in Table 5 in mitigating the reverse causality concern.

Table C1: Fixed-effect panel models for alternative measures of digital proximity, with log(q) as the dependent variable. Models (1) to (3) test count-based digital proximity measures, and Models (4) to (6) test sales-based digital proximity measures. In Models (1) and (4) the digital proximity measure is the shortest path to/from IT in an undirected industry space. In Models (2) and (5) the digital proximity measure is the shortest path to IT, and in Models (3) and (6) the measure is the shortest path from IT. These measures are all developed based on a directed industry space network. Huber-White robust errors are clustered on the firm identifiers. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Variable	(1) Count-based/Undirected	(2) Count-based/to IT	(3) Count-based/from IT	(4) Sales-based/Undirected	(5) Sales-based/to IT	(6) Sales-based/from IT
Digital Proximity	0.180***	0.115*	0.173***	0.179***	0.180***	0.182***
	(0.0583)	(0.0645)	(0.0631)	(0.0607)	(0.0607)	(0.0622)
Constant	1.426***	1.458***	1.449***	1.426***	1.426***	1.427***
	(0.190)	(0.192)	(0.193)	(0.191)	(0.191)	(0.190)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year and Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	22,280	22,280	22,280	22,278	22,278	22,277
R-squared	0.151	0.151	0.151	0.151	0.151	0.151
Unique Firms	3,794	3,794	3,794	3,793	3,793	3,793
F stat	51.97***	51.93***	52***	52.22***	52.22***	52.21***

Table C2: Effect of digital proximity on the logarithm of Tobin's q. We compare fixed-effects (FE) in Model (1), random-effects (RE) in Model (2), generalized estimation equations (GEE) with exchangeable within group correlations in Model (3), and autoregressive within group (AR1) correlations in Model (4). Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Variable	(1) FE	(2) RE	(3) GEE	(4) GEE
Digital Proximity	0.180***	0.140***	0.138***	0.115**
	(0.0583)	(0.0495)	(0.0488)	(0.0560)
log(Ind. Avg. Tobin's q)	0.0157***	0.0249***	0.0264***	0.0139***
	(0.00431)	(0.00409)	(0.00409)	(0.00388)
R&D Intensity	0.0529***	0.118***	0.126***	0.104***
	(0.0196)	(0.0158)	(0.0155)	(0.0158)
Advertising Intensity	0.254	0.655***	0.710***	0.694***
	(0.243)	(0.192)	(0.187)	(0.186)
Ind. Avg. Closeness Centrality	0.732	0.215	0.173	1.724
	(1.743)	(1.354)	(1.336)	(1.342)
Intangible Assets	-5.47e-05**	-1.53e-05	-7.87e-06	5.72e-05**
	(2.74e-05)	(2.35e-05)	(2.32e-05)	(2.84e-05)
log(Total Assets)	-0.170***	-0.114***	-0.111***	-0.111***
	(0.0121)	(0.00593)	(0.00572)	(0.00656)
Diversification	-0.0685**	-0.0881***	-0.0911***	-0.0366
	(0.0309)	(0.0258)	(0.0253)	(0.0306)
HHI	-0.0586	-0.116***	-0.122***	-0.121***
	(0.0389)	(0.0319)	(0.0313)	(0.0339)
Weighted Market Share	0.107	0.220***	0.232***	0.235***
	(0.0789)	(0.0669)	(0.0657)	(0.0642)
Constant	1.426***	1.142***	1.125***	1.358***
	(0.190)	(0.184)	(0.184)	(0.173)
Observations	22,280	22,280	22,280	13,943
R-squared	0.151	-	-	-
Unique Firms	3,794	3,794	3,794	2,384
F stat/Wald chi2(df)	51.97*** (37)	2714*** (38)	2806***	1877***

Table C3: Effect of digital proximity on past, present, and future values of Tobin's q, with fixed-effects panel regression models. Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Variable	(1) l5.log(q)	(2) l4.log(q)	(3) l3.log(q)	(4) l2.log(q)	(5) l1.log(q)	(6) Present q	(7) f1.log(q)	(8) f2.log(q)	(9) f3.log(q)	(10) f4.log(q)	(11) f5.log(q)
Digital Proximity	-0.0642	-0.0801	0.0644	0.0556	0.0639	0.191***	0.165***	0.126*	0.135*	0.0818	0.0271
	(0.0833)	(0.0737)	(0.0670)	(0.0634)	(0.0602)	(0.0580)	(0.0607)	(0.0647)	(0.0732)	(0.0734)	(0.0703)
log(Ind. Avg. Tobin's q)	-0.00736	-0.0138***	0.00398	0.0106**	0.00897*	0.0156***	0.0119***	-0.00677	-0.00470	-0.0102*	-0.0186***
	(0.0049)	(0.0048)	(0.0047)	(0.0048)	(0.0046)	(0.0043)	(0.0044)	(0.0046)	(0.0049)	(0.0052)	(0.0053)
R&D Intensity	-0.0180	0.0132	0.0356	0.0555**	0.0534**	0.0528***	0.0407*	0.0152	0.0407	0.0284	0.0489
	(0.0319)	(0.0306)	(0.0267)	(0.0255)	(0.0255)	(0.0196)	(0.0218)	(0.0248)	(0.0254)	(0.0278)	(0.0310)
Advertising Intensity	-0.479	-0.0758	0.562*	0.754**	0.596**	0.252	-0.319	0.170	0.330	0.287	0.275
	(0.413)	(0.372)	(0.329)	(0.297)	(0.264)	(0.243)	(0.244)	(0.253)	(0.274)	(0.291)	(0.291)
Ind. Avg. Closeness Centrality	-1.492	-1.608	-1.357	-2.617	-0.897	0.668	-0.835	-0.301	-0.599	0.504	-0.828
	(2.524)	(2.473)	(2.365)	(2.068)	(1.836)	(1.744)	(1.706)	(1.748)	(1.772)	(1.737)	(1.750)
Intangible Assets	-1.63e-05	-3.24e-05	-5.78e-05*	-7.22e-05**	-7.89e-05***	-5.40e-05**	2.09e-06	-3.94e-05	-7.11e-05**	-8.77e-05***	-8.40e-05**
	(3.85e-05)	(3.66e-05)	(3.28e-05)	(2.89e-05)	(2.79e-05)	(2.75e-05)	(2.77e-05)	(2.80e-05)	(2.98e-05)	(3.17e-05)	(3.51e-05)
log(Total Assets)	0.0833***	0.0859***	0.0734***	0.0555**	-0.0109	-0.169***	-0.227***	-0.186***	-0.134***	-0.0894***	-0.0647***
	(0.0169)	(0.0157)	(0.0145)	(0.0139)	(0.0135)	(0.0121)	(0.0108)	(0.0106)	(0.0105)	(0.0110)	(0.0120)
Diversification	-0.0742	-0.107**	-0.0873**	-0.0999***	-0.111***	-0.0434	-0.0165	0.0165	0.0465	0.0521	0.0433
	(0.0459)	(0.0426)	(0.0393)	(0.0363)	(0.0342)	(0.0320)	(0.0319)	(0.0323)	(0.0329)	(0.0340)	(0.0363)
HHI	-0.0112	-0.0229	-0.0442	-0.0279	-0.0881**	-0.0590	-0.0502	-0.0245	-0.0125	-0.0254	-0.0142
	(0.0472)	(0.0481)	(0.0474)	(0.0458)	(0.0424)	(0.0390)	(0.0399)	(0.0412)	(0.0429)	(0.0448)	(0.0497)
Weighted Market Share	0.164*	0.142	0.107	0.0750	0.0759	0.107	0.0648	-0.000990	0.000450	0.0357	0.0624
	(0.0918)	(0.0937)	(0.0921)	(0.0865)	(0.0820)	(0.0788)	(0.0832)	(0.0829)	(0.0832)	(0.0848)	(0.0938)
Constant	0.221	0.386**	0.358**	0.141	0.573***	1.424***	1.944***	1.730***	1.466***	1.229***	1.165***
	(0.235)	(0.166)	(0.167)	(0.172)	(0.184)	(0.190)	(0.124)	(0.134)	(0.149)	(0.143)	(0.147)
Firm and Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	13,928	15,183	16,707	18,468	20,426	22,280	20,956	19,208	17,166	15,281	13,556
R-squared	0.112	0.112	0.110	0.113	0.105	0.152	0.193	0.164	0.133	0.108	0.095
Number of Unique Firms	2,177	2,465	2,799	3,149	3,513	3,794	3,673	3,437	3,134	2,860	2,577
F stat	31.70***	35.99***	34.92***	38.42***	39.43***	52.03***	51.86***	45.16***	37.66***	29.90***	25.90***

Table C4: Effect of Tobin's q on past, present, and future values of digital proximity, with fixed-effects panel regression models. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Variable	(1) Pres. Dig. Prox.	(2) Pres. Dig. Prox.	(3) Pres. Dig. Prox.	(4) Pres. Dig. Prox.	(5) Nxt Yr Dig. Prox.	(6) Nxt Yr Dig. Prox.	(7) Nxt Yr Dig. Prox.
log(Tobin's q)	0.00301	0.00312			0.00341		
	(0.00252)	(0.00200)			(0.00222)		
1st lag log(Tobin's q)	-0.000412		0.00182			0.000998	
	(0.00263)		(0.00191)			(0.00255)	
2nd lag log(Tobin's q)	7.81e-05			0.00128			0.00175
	(0.00246)			(0.00222)			(0.00217)
3rd lag log(Tobin's q)	0.00231						
	(0.00228)						
4th lag log(Tobin's q)	-0.00400*						
	(0.00224)						
5th lag log(Tobin's q)	0.00208						
	(0.00241)						
6th lag log(Tobin's q)	0.00166						
	(0.00217)						
Dig. Proximity					0.544***	0.547***	0.545***
					(0.0297)	(0.0296)	(0.0297)
1st lag Dig. Proximity	0.485***	0.500***	0.499***	0.500***	0.00538	0.00455	0.00524
	(0.0294)	(0.0279)	(0.0279)	(0.0279)	(0.0220)	(0.0220)	(0.0221)
2nd lag Dig. Proximity	0.00455	0.0146	0.0157	0.0145	-0.0330	-0.0317	-0.0322
	(0.0219)	(0.0207)	(0.0208)	(0.0208)	(0.0201)	(0.0202)	(0.0201)
3rd lag Dig. Proximity	-0.0178	-0.0276	-0.0281	-0.0267	0.0124	0.0125	0.0122
	(0.0145)	(0.0172)	(0.0172)	(0.0173)	(0.0246)	(0.0247)	(0.0248)
4th lag Dig. Proximity	-0.000343	0.00543	0.00434	0.00437	0.0232	0.0232	0.0224
	(0.0244)	(0.0231)	(0.0231)	(0.0232)	(0.0195)	(0.0195)	(0.0195)
5th lag Dig. Proximity	0.0169	0.0136	0.0141	0.0136	-0.0382**	-0.0381**	-0.0378**
	(0.0189)	(0.0171)	(0.0171)	(0.0171)	(0.0159)	(0.0159)	(0.0159)
6th lag Dig. Proximity	-0.0605***	-0.0548***	-0.0537***	-0.0543***	-0.0152	-0.0152	-0.0155
	(0.0177)	(0.0160)	(0.0160)	(0.0160)	(0.0220)	(0.0220)	(0.0220)
Constant	0.272***	0.234***	0.239***	0.238***	0.132***	0.137***	0.138***
	(0.0202)	(0.0295)	(0.0295)	(0.0296)	(0.0332)	(0.0329)	(0.0330)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm and Year Fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	11,174	12,286	12,286	12,266	10,710	10,708	10,688
R-squared	0.382	0.391	0.391	0.391	0.341	0.343	0.341
Unique Firms	1,749	1,935	1,938	1,934	1,726	1,728	1,722
F stat	21310***	42.04***	40.69***	40.76***	38.99***	39.18***	38.57***

References

Chang, Y. B., and Gurbaxani, V. 2012. "Information Technology Outsourcing, Knowledge Transfer, and Firm Productivity: An Empirical Analysis," *MIS Quarterly* (36:4), pp. 1043-1063 (doi: 10.2307/41703497).

Farjoun, M. 1994. "Beyond Industry Boundaries: Human Expertise, Diversification and Resource-related Industry Groups," *Organization Science* (5:2), pp. 185-199.

Farjoun, M. 1998. "The Independent and Joint Effects of the Skill and Physical Bases of Relatedness in Diversification," *Strategic Management Journal* (19:7), pp. 611-630 (doi: 10.1002/(SICI)1097-0266(199807)19:7<611::AID-SMJ962>3.0.CO;2-E).

Greene, W. H. (2003). Econometric analysis. Pearson Education India.

Hidalgo, C., and Hausmann, R. 2009. "The Building Blocks of Economic Complexity," *Proceedings of the National Academy of Sciences* (106:26), pp. 10570-10575 (doi: 10.1073/pnas.0900943106).

60

Hidalgo, C. A., Klinger, B., Barabási, A.-L., and Hausmann, R. 2007. "The Product Space Conditions the Development of Nations," *Science* (317:5837), pp. 482-487 (doi: 10.1126/science.1144581).

Im, K. S., Dow, K. E., and Grover, V. 2001. "Research Report: A Reexamination of IT Investment and the Market Value of the Firm—An Event Study Methodology," *Information Systems Research* (12:1), pp. 103-117.

Kleinberg, J. M. 1999. "Authoritative Sources in a Hyperlinked Environment," *Journal of the ACM (JACM)*, (46:5), pp. 604-632.

Kleis, L., Chwelos, P., Ramirez, R. V., and Cockburn, I. 2012. "Information Technology and Intangible Output: The Impact of IT Investment on Innovation Productivity," *Information Systems Research* (23:1), pp. 42-59 (doi: 10.1287/isre.1100.0338).

Kogan, L., Papanikolaou, D., Seru, A., and Stoffman, N. 2017. "Technological Innovation, Resource Allocation, and Growth," *The Quarterly Journal of Economics* (132:2), pp. 665-712 (doi: 10.1093/qje/qjw040).

Lee, C. H., Venkatraman, N., Tanriverdi, H., and Iyer, B. 2010. "Complementarity-based Hypercompetition in the Software Industry: Theory and Empirical Test, 1990–2002," *Strategic Management Journal* (31:13), pp. 1431-1456 (doi: 10.1002/smj.895).

Milgrom, P., and Roberts, J. 1990. "The Economics of Modern Manufacturing: Technology, Strategy, and Organization," *American Economic Review* (80:3), pp. 511-528.

Porter, M. 2008. *Competitive Strategy: Techniques for Analyzing Industries and Competitors*, New York, NY: Simon and Schuster.

Porter, M., and Siggelkow, N. 2008. "Contextuality within Activity Systems and Sustainability of Competitive Advantage," *Academy of Management Perspectives* (22:2), pp. 34-56 (doi: 10.5465/AMP.2008.32739758).

Tanriverdi, H., Rai, A., and Venkatraman, N. 2010. "Research Commentary—Reframing the Dominant Quests of Information Systems Strategy Research for Complex Adaptive Business Systems," *Information Systems Research* (21:4), pp. 822-834.