# Digital technology adoption and knowledge flows within firms: Can the Internet overcome geographic and technological distance?☆

Chris Forman[a], Nicolas van Zeebroeck[b,*]

[a] Dyson School of Applied Economics and Management, Cornell University, 137 Reservoir Avenue, Ithaca, NY 14853, USA
[b] iCITE, Solvay Brussels School of Economics and Management, Université libre de Bruxelles, 50, avenue F.D. Roosevelt - CP114, 1050 Brussels, Belgium

## ARTICLE INFO

*Keywords:*
Geography of innovation
Internet adoption
IT investments
Knowledge spillovers
Patent citations
Technological proximity
Technological specialization

## ABSTRACT

Under what conditions does digital technology adoption increase cross location knowledge flows within firms? We investigate this question by studying the impact of adopting basic Internet access on cross-location knowledge flows within the same firm. We construct a large data set of Internet adoption and patent citations among dyadic pairs of firm-locations between 1992–1998. We find that when both locations in the pair adopt basic Internet there is an increase in the likelihood of a citation between the citing and (potential) cited location. In contrast, we find no significant effect of Internet adoption at only the citing location. We further study how this effect varies according to the proximity of the research activities between the source and recipient of knowledge and specialization of the research activities within the recipient. We find that the likelihood of a citation increases more after dyadic Internet adoption when the pair is working in similar research areas and when the research areas in the citing location are less specialized. These results, which are robust to a range of robustness analyses, suggest that digital technologies such as Internet connectivity are able to facilitate knowledge flows between locations only when they share a common knowledge base.

## 1. Introduction

Research and development operations are often physically dispersed within firms (Leiponen and Helfat, 2011; Miller et al., 2007; Penner-Hahn and Shaver, 2005; Singh, 2008). Such geographic dispersion confers benefits, including the ability to tap into local knowledge bases. However, even within firms, it is often difficult to effectively transfer and use knowledge produced elsewhere within an organization (Allen, 1977; Teece, 1977). The increasing digitization of innovation processes has the potential to increase the flows of knowledge across locations within geographically disparate organizations. However, there are often significant barriers to the transfer of knowledge through digital systems.[1] While the sources and consequences of these barriers have been explored in the literature, a lack of systematic empirical evidence has made it difficult to quantify their effects and assess the conditions in which digital technology can help overcome the well-known localization of knowledge (e.g. Jaffe et al., 1993), leaving open a number of important questions.

One question of particular interest is whether digital technology helps users to identify and absorb knowledge that they do not already know, or simply reinforces existing areas of expertise. One view has held that IT systems can enable users to access new sources of knowledge outside of their own local cohort or group (e.g., Offsey, 1997), leading to novel recombinations of ideas that generate significant scientific impact (Uzzi et al., 2013) and that are important for economic growth (Weitzman, 1998). Another has held that individuals use IT systems to disproportionately reach out to existing contacts and those who share similar tastes or knowledge bases (Gaspar and Glaeser, 1988; Rosenblat and Mobius, 2004; Van Alstyne and Brynjolfsson, 1996, 2005). Thus, there is the potential for adoption of digital technology[2] to reduce the importance of geographic distance but at the same time increase distance

[1] See Alavi and Leidner (2001) for one review.

[2] In this paper we use the terms "digital technology", "Information Technology (IT)" and "Information and Communication Technology (ICT)" as synonyms. Our focus is on "basic Internet technology", which primarily involves World Wide Web (including intranets) access and SMTP-based email. "Digitization" refers to the use of digital technology to convert information and processes into digital artifacts. We define "digital innovation" as in Nambisan et al. (2017): "the creation of (and consequent change in) market offerings, business processes, or models that results from the use of digital technology."

arising from differences in technical expertise (Van Alstyne and Brynjolfsson, 2005).

To understand the implications of digital technology adoption for reducing barriers arising from technological and geographic distance, we begin by constructing a framework for understanding how digital technology adoption influences R&D-related knowledge flows. This framework documents differences in the barriers to IT-enabled knowledge flows arising from the nature of technology and the presence of collaborations between scientists in the two locations. Once we have developed this framework, we examine the implications of the rapid declines in communication costs that occurred around the time of the initial commercialization of the Internet. Our research approach combines detailed data related to Internet adoption with that on patent citation patterns within large organizations.

We first examine the average change in patent citations that occurs between dyadic pairs of firm locations after one or both locations adopt Internet technology in the late 1990s. Because the process through which Internet technology adoption will influence knowledge flows will depend upon whether scientists are collaborating on a joint research project, we examine separately the cases in which patents do and do not arise from a collaboration. We show that *dyadic Internet adoption*—that is, technology adoption at both citing and cited locations within the firm—leads to an increase in the likelihood of a citation from patents that do not arise from a collaboration between the two locations of 1.2 percentage points, or 17.6%. These results are stronger statistically and economically than those arising from technology adoption at only the citing location; controlling for the effects of dyadic technology adoption, adoption at the citing location only has no statistically significant effect on citation behavior. We present evidence from a range of robustness analyses to probe whether these results likely reflect a causal relationship. In particular, we provide results from regressions in which we instrument for Internet adoption using variables that will influence the costs of adopting Internet technology but which should have little impact on knowledge flows within firms.

We next turn to the primary goal of the paper—to explore whether there are differential effects of technology adoption on knowledge flows from close and distant technological areas. We find evidence that adoption of Internet technology leads to knowledge flows that reinforce existing areas of expertise. In particular, we provide evidence that adoption of dyadic Internet increases the likelihood of knowledge flows across all types of patents (i.e., those both arising and not arising from a collaboration) only when the two locations are close in technological distance or when the citing location is not technologically specialized across a small number of fields. Adoption at the citing location only has little effect on knowledge flows, whether they are specialized or not. In short, while our results provide evidence of technology adoption leading to an increase in geographically distant knowledge flows, this holds only for cases in which the citing and cited locations share common technological areas.

## 1.1. Related literature

Our findings contribute to several fields of research. First, we relate to a broad literature that has demonstrated that knowledge flows are localized (e.g., Jaffe et al., 1993; Thompson and Fox-Kean, 2005).[3] In particular, we are most closely related to studies that have shown the benefits of co-location to knowledge exchange and innovation within organizations (Allen, 1977), and that has demonstrated the impact of co-location on innovation, commercialization, and product quality outcomes (e.g., Catalini, 2018; Bercovitz and Feldman, 2011; Gray et al., 2015). We also relate to a line of work that investigates the efficacy of organizational practices that shape knowledge flows and knowledge exchange within geographically dispersed research

organizations (e.g., Hansen, 1999; Henderson and Cockburn, 1994; Zhao, 2006). However, these works have not illustrated the impact of digital technology adoption on knowledge flows.

Our research is also related to a line of work that has investigated whether IT investments have changed the geography of scientific production.[4][5] Early network technologies such as BITNET increased scientific collaboration, specifically across multiple universities (Agrawal and Goldfarb, 2008; Ding et al., 2010; Walsh et al., 2000). However, research that has investigated the effects of digital technology on the benefits of co-location within firms has thus far delivered somewhat mixed results. In a paper that is perhaps closest to this one, Forman and van Zeebroeck (2012) use a similar cross-industry data set to provide evidence that adoption of an early generation of Internet technology lead to an increase in collaborations between researchers located in distant geographic units within firm. However, other work has shown that Internet adoption may not reduce the benefits of co-location between manufacturing and R&D (Gray et al., 2015).

We also speak to recent work that has argued that while new communication technologies increase collaboration and knowledge flows across physical space, they may also reduce interaction between individuals with differing tastes, preferences, and social/professional circles (Rosenblat and Mobius, 2004; Van Alstyne and Bryjolfsson, 2005). Some have argued that this may result in the balkanization of science (Van Alstyne and Brynjolfsson, 1996). While we are unable to confirm the specific data-generating processes argued for in these earlier papers, we do provide empirical evidence that supports the view that through lower communication costs, digital technology will differentially increase knowledge flows from researchers of similar fields relative to those who work in different research areas.

## 2. Research framework

Geographic distance has been shown to be a barrier to knowledge flows within and between firms (Allen, 1977; Teece, 1977). While digital technology reduces communication costs, prior work has suggested that there may be considerable barriers to IT-enabled knowledge transfer (e.g., Powell, 1998; Anderson et al., 2018).

Knowledge flows over geographical distance have been widely studied in the field of economics of innovation. Our empirical tests are motivated by this research, but also by the related field of knowledge management and more specifically the knowledge-based view (KBV) of the firm (Spender, 1996; Grant, 1996). In this stream of work, the firm is viewed as a set of institutional arrangements to integrate and apply knowledge at different levels (Grant, 1996). The KBV extends beyond the mere transfer of knowledge to encompass its reuse or application elsewhere to create new knowledge. While digital technology has been studied as a potentially powerful channel to achieve knowledge integration, there has been limited systematic empirical evidence of the kind that we present on the impact of specific digital technologies on knowledge integration.

More importantly, knowledge integration is known to be harder to achieve across organizational boundaries defined by distance, technological field, or other dimensions (e.g., Levina and Vaast, 2005; Szulanski, 1996; Teece, 1977). The main reason lies in the tacit nature of specialized knowledge (e.g., Polanyi, 1958). Knowledge is tacit when "*the aim of skillful performance is achieved by the observance of a set of*

---

[3] For a recent review of this literature, see Feldman and Kogler (2010).

[4] Another line of works has investigated whether IT investment is associated with an increase in innovation and productivity more broadly (e.g., Hall et al., 2013; Kleis et al., 2012; Ravichandran et al., 2017; Wu et al., 2016).

[5] We further acknowledge a significant literature in the field of information systems on knowledge management (for one review see Alavi and Leidner, 2001). While this research informs our thinking of the challenges of knowledge absorption, it does not centrally focus on the issues of digital innovation that are our central concern.

| | | CITING PATENT RESULTS FROM A COLLABORATION | |
|---|---|---|---|
| | | NO | YES |
| **ADOPTION** | **DYADIC** | (A) Facilitates person-to-person interaction and access to documents (STRONG) | (C) Facilitates person-to-person interaction (through creation of a collaboration and likelihood of transfer conditional on collaboration) and access to documents (STRONG) |
| | **CITING ONLY** | (B) Facilitates access to documents (WEAK) | (D) Facilitates access to documents (WEAK) |

**Fig. 1.** Expected effects of Internet adoption based on adoption type and collaborative nature of citing patents. A patent is deemed to result from a collaboration between locations A and B if there is at least one inventor from each location listed as an inventor of the focal patent.

*rules which are not known as such to the person following them.*" (Polanyi, 1958, p. 49) Transferring such tacit knowledge requires its expression (converting tacit knowledge into explicit knowledge, 'articulation' in the words of Nonaka (1991)) and absorption. This in turn necessitates the development of common codes and knowledge bases across the source and recipient of tacit knowledge (Nonaka, 1991; Nonaka and Takeuchi, 1995).

Nonaka (1991) denotes as "combination" the transfer of explicit knowledge from one specialized individual or group to another. Arguably, this is the type of transfer that would benefit the most from digitization of documents since the knowledge is readily available in explicit (hence transferrable) form: digitization facilitates communication of existing explicit knowledge. In contrast, applying and reusing tacit knowledge requires tacit to explicit (articulation) conversions (Nonaka, 1991). By and large, the knowledge management literature points at direct interpersonal interactions as the key channels for such transfers, typically in the form of collaborative work (e.g. cross-functional teams) or job rotation (Nonaka and Takeuchi, 1995; Moenaert and Souder, 1990). By lowering communication costs of interpersonal interactions through technologies like email, the digital technologies we study may also facilitate this type of knowledge transfer. However, prior literature has suggested that digital technology may be less effective at these types of transfers than when transferring explicit knowledge (Alavi and Leidner, 2001).

In sum, there is potential variance in the ability of digital technology to support knowledge transfers depending on the type of transfer involved and in the type of transmission channel used.

### 2.1. Internet adoption and knowledge flows within an R&D organization

We examine the impact of Internet adoption on knowledge flows within an R&D organization. Our research approach is informed by our setting. We follow a long line of prior literature starting with Jaffe, Trajtenberg, and Henderson (1993) and use citations between patents to measure knowledge flows. It is worth stressing that a citation reflects not just a transfer of knowledge from the source (the inventors of the cited patent) to the recipient (the inventors of the citing patent) but an actual instance of reuse or application as proved by the new (citing) patent. A citation is therefore deemed to reflect the actual diffusion and reuse of knowledge (i.e. integration) and forms our operational definition of a knowledge flow.

Our focus is on the adoption of Internet technology around the time of the initial commercialization of the Internet. In our baseline estimation strategy we examine the implications of Internet technology adoption on the likelihood of a citation between two geographically dispersed locations within an organization: that is, we estimate the change in the incidence of citations between any two geographically-distant establishments within a given firm when one or both of the firm locations adopt Internet, compared to when the locations do not adopt.

In our setting, there are two aspects of how the technology is deployed that can inform the mechanisms through which it contributes to knowledge flows.

First, depending upon how it is adopted, Internet technology can facilitate two different types of search processes. Internet adoption can aid direct communication between users within an organization, as when an inventor asks a colleague for help (e.g., Allen, 1977). It can also lower the costs of searching for now-digitized documents. As noted above, direct communication may be more effective than static documents at facilitating transfer of tacit knowledge. While our data do not allow us to directly observe the process of information discovery within an organization, our research design will enable us to separately identify the effects of dyadic Internet adoption from that where only the citing location adopts. We will use this distinction to provide some insights into the relative importance of these different processes.

Second, within our setting we will also separately consider citations that occur between two locations as a result of a patent arising from a collaboration, and citations between the two locations that do not arise as the result of a collaboration (i.e., there is no inventor based at the cited location listed on the citing patent). This distinction may be significant since in the context of a collaborative patent Internet connectivity may come in support of the type of person-to-person interaction that is more effective at facilitating transfer of tacit knowledge. The underlying mechanism through which digital technology contributes to knowledge flows may therefore be different between collaborative and non-collaborative work.

Fig. 1, summarizes these two dimensions with our predictions on the magnitude of the effect of Internet adoption on knowledge flows, which we discuss next:

### 2.1.1. Knowledge flows arising from adopting digital technology at both locations outside of collaborative work

In this subsection we investigate the effects of dyadic Internet adoption on knowledge flows through direct interaction with distant co-workers outside of a collaboration (cell A in Fig. 1). The process for knowledge transfer from other employees can be motivated by models of search (Garicano, 2000). In our setting, a piece of knowledge found elsewhere in the organization may help the employee to solve a problem that arises within the research process. Employees continue to search so long as the benefits of search exceed the costs. As they continue to search, they may be able to identify a better solution to their problem. However, search is costly, and the outcome of future search is also uncertain. For one, additional search may not yield a new idea that is useful. Further, even if employees hear a solution to their problem they may not be able to apply it effectively if they do not have the requisite knowledge (Cohen and Levinthal, 1989, 1990; Nelson and Winter, 1982). This directly points at the difference between knowledge diffusion and knowledge integration.

Adoption of new Internet technology reduces the search costs of

accessing knowledge. It will also reduce transfer costs and facilitate knowledge absorption ("articulation" in the words of Nonaka (1991)), if Internet technology reduces the costs of person-to-person communication (e.g., through email). If the combination of search and transfer costs fall sufficiently, then this increases the likelihood of a transfer of knowledge. This represents a joint hypothesis that adoption of Internet technology by both the sender and receiver will increase the likelihood of search as well as the likelihood of knowledge flow conditional on search.

### 2.1.2. Knowledge flows arising from adopting digital technology at one location outside of collaborative work

We next consider the effects of Internet adoption at only the citing (source) location on knowledge flows outside of a collaboration (cell B in Fig. 1). Knowledge can be transferred in an organization through the production and reading of technical documents (Allen, 1977). These can include formal documents that are intended to be published externally, such as patents or journal publications, or purely internal documents such as lab notes. Like the transfer of knowledge arising through dyadic communication among people, the transfer of knowledge through technical documents can be viewed through a model of search. Users will continue searching for documents until the benefits no longer exceed costs. In short, adoption of Internet technology reduces the costs of transferring information that has already been codified, or explicit-to-explicit transfers ("combination" in the words of Nonaka (1991)).

However, unlike the mechanisms described in the prior section, use of Internet technology in this way will not facilitate digitization of ad hoc interactions that arise from questions about the underlying knowledge and how it may improve upon and interact with the components of an existing invention under development. In other words, adoption at only one side of the location-pair dyad will not facilitate the transfer of tacit knowledge, hence tacit to explicit conversions.

### 2.1.3. Digital technology adoption and knowledge transfer that arises from a collaboration

Cells C and D of Fig. 1 present predictions of the effects of digital technology on citations mediated through collaborations. Internet technology can influence the likelihood of a citation mediated through collaborations in two ways: by increasing the likelihood of a collaboration and by increasing the likelihood of a citation conditional on a collaboration.

Prior work has demonstrated that adoption of digital technology reduces the costs to geographically distant collaborations, and so will increase the incidence of these types of collaborations (Agrawal and Goldfarb, 2008; Ding et al., 2010; Forman and van Zeebroeck, 2012). By creating a new team that spans locations, adoption of IT systems will increase the likelihood of drawing on knowledge from both locations in the pair. Firms with strong internal linkages such as cross-location teams will be more effective at drawing on complementary knowledge and resources within the firm (Alcacer and Zhao, 2012).

Thus, we expect that IT systems will also increase the likelihood of knowledge flows between locations by increasing the incidence of collaborations between locations and by facilitating knowledge transfers within the context of an active collaboration. However, IT-enabled collaborations will require adoption of digital technology across both locations in the pair: pairs in which only the receiving location adopts will not see an increase in collaborations (Agrawal and Goldfarb, 2008; Forman and van Zeebroeck, 2012).

Adoption of digital technology can also potentially increase the likelihood of knowledge flows conditional on a collaboration. In our research design we will be unable to separate the effects of technology adoption on the likelihood of a collaboration and the likelihood of knowledge flow conditional on a collaboration. Rather, we observe the combined effects of these two processes.

### 2.2. Conditions when the effects of digital technology adoption on knowledge flows will be strongest

Novel combinations of new ideas have been shown to lead to the most important discoveries (Weitzman, 1998; Boudreau et al., 2011; Jones, 2009; Uzzi et al., 2013). Despite the potential benefits that can be derived from using digital technology to access new ideas from dissimilar fields, it is well known that searching for and combining knowledge from dissimilar fields can be challenging (Fleming, 2001; Kogut and Zander, 1992; Schilling and Green, 2011). As noted above, digital technology may have uneven efficacy of transferring different types of knowledge. Because of the differing benefits of integrating different types of knowledge and uneven benefits of Internet adoption in facilitating knowledge flows, it is unclear ex ante whether adoption of Internet technology will facilitate flows of similar or dissimilar knowledge.

In the next section, we describe our efforts to take a first step at answering this question. We do this by measuring how variance in the concentration of prior knowledge at the recipient location and proximity of prior knowledge of inventors between the source and recipient influences the manner in which Internet technology adoption influences cross-location knowledge flows. We first discuss our baseline case of non-collaborative patents, and then discuss the case of collaborative patents.

### 2.2.1. Moderating role of technological proximity and specialization outside of collaborations

Searching across different research domains is generally more costly and uncertain relative to searching within domains (Fleming, 2001; Schilling and Green, 2011). In addition, inventors in related fields may have stronger social ties that have been developed through prior collaborations or interactions. As noted above, the types of digital technologies that we study can facilitate access to documents through the web or facilitate direct communication with pre-existing sources or existing social connections. However, it is not as effective at establishing new social connections (Gaspar and Glaeser, 1998). It is well known that searching across organizational units may be difficult because of a lack of direct communication and relationships (e.g., Allen, 1977; Galbraith, 1973), and this absence of prior relationships may make it difficult for knowledge transfer even when the costs of communication are lowered by digital technology. For cases when only the citing location adopts, users may not be aware of databases or intranet pages where the knowledge resides (Jordan and Jones, 1997).

Further, even conditional on identifying useful knowledge, inventors may experience challenges absorbing and putting to use new ideas from different areas (Cohen and Levinthal, 1989, 1990; Henderson and Clark, 1990). As discussed above, such challenges may have different causes.[6] This points again at Nonaka's notion of "articulation" that is necessary to achieve integration.

For these reasons and others, several authors have claimed that face-to-face interactions and worker mobility are better channels than digital technology to transfer knowledge across fields or functions (Gupta and Govindarajan, 2000; Kogut and Zander, 1992). Since dyadic Internet adoption facilitates bilateral interactions, it may have an advantage in facilitating transfer across fields relative to adoption at only the citing location.

In sum, other things equal the (gross) benefits of Internet-enabled knowledge transfer will be greatest among dissimilar or more specialized knowledge bases for which the benefits of use are greatest. However, the margin of Internet technology that we study will be most effective in helping inventors to search for and transfer knowledge in similar fields. Because proximity in research will influence costs and

---

[6] For examples, see Kogut and Zander (1992); Leonard-Barton (1988), and Fleming (2001).

benefits in opposite directions, we are unable to sign ex ante the net effects of these differences, and instead use the data to determine them.

### 2.2.2. Moderating role of technological proximity and specialization in the context of collaborations

As noted above, we expect Internet adoption to contribute both to the incidence of collaborative work and to the intensity of knowledge transfer within an active collaboration. This effect may also be influenced by the heterogeneity and concentration of the knowledge bases at the two participating locations. As noted before, the benefits of combining knowledge from unrelated fields is high. However, collaborations are seen as one way of facilitating recombination across distant fields (Jones, 2009) and the primary mechanism to achieve cross-functional integration (Moenaert and Souder, 1990). Therefore, the implication of Internet adoption might be different in the context of collaborative research projects.

While the benefits of collaborations across fields may be higher, collaborations between workers in different specialized areas is also more difficult (Becker and Murphy, 1992). Prior work has demonstrated the ability of digital communication technology to reduce the costs of collaborating across physical distance (Agrawal and Goldfarb, 2008; Forman and van Zeebroeck, 2012), but has also shown there are significant costs of technology-enabled distributed work when workers do not have a shared knowledge base (Srikanth and Puranam, 2011). Thus, adoption of Internet technology may be unable to facilitate collaborations across distant and/or narrower fields.

Provided such a collaboration nonetheless occurs, Internet technology will – per our earlier discussion – likely be more efficient at transferring and integrating knowledge within a given field than across fields. Consequently, within collaborations as well as outside of collaborations, the effect of dyadic Internet adoption on knowledge flows may increase when scientists are working in similar or overlapping research areas.

### 3. Empirical strategy

#### 3.1. Estimating the average effect of Internet on citations

In our baseline estimation approach we used fixed effects panel data regression models. The unit of observation is a within-firm location dyad-year, and due to data constraints we observe behavior every other year. Our use of fixed effects will difference out the effects of time-invariant dyad unobservables that may influence the incidence of citations between the two establishments. Our approach yields the following estimating equation:

$$Citation_{ijkt} = \alpha_1 X_{ijkt} + \alpha_2 Z_{ijkt} di + \beta Internet_{ijkt} + \mu_{ijk} + \tau_t + \varepsilon_{ijkt} \qquad (1)$$

$Citation_{ijkt}$ indicates the existence of at least one US patent of firm $i$ that was applied for in year $t$ or the preceding year, that was invented in location $j$, and that cites another patent invented in location $k$ of the same firm in the 10 preceding years (we have explored robustness to alternative time windows). Depending upon the specification, $Citation_{ijkt}$ can represent citations that arise from patents that are the result of a collaboration between locations $j$ and $k$ in the citing patent (which we label *collaborative patents*), patents that do not arise as a result of such a collaboration (which we label *non-collaborative patents*), or citations that arise from either collaborative or non-collaborative patents. Our main variable of interest is $Internet_{ijkt}$, which indicates whether Internet had been adopted. We consider two types of Internet adoption: *dyadic adoption* exists when both locations $j$ and $k$ of firm $i$ had both adopted Internet in or before year $t$, while *citing only adoption* exists as soon as the citing location $j$ adopts has adopted by year $t$.

$X_{ijkt}$ is a vector of time-varying controls at the establishment-pair level such as the log of patent stock in the pair over the prior 10-year period and the log of per-establishment R&D spending. $Z_{ijkt}$ is a vector

of time-varying controls for local characteristics such as local average weekly wages and the log of local employment. $\mu_{ijk}$ measures pair fixed effects and $\tau_t$ measures time fixed effects.

We utilize the fact that in the first two years of our data, 1992 and 1994, adoption of basic Internet within firms in our sample will be equal to zero. This is because these two years predate the commercialization of the Internet. For the parameters in Eq. (1) to be identified, we require significant within-firm variance in basic Internet adoption within firms in 1996 and 1998, over the period when the commercial Internet began to diffuse. Using a similar sample, Forman and van Zeebroeck (2012) demonstrate significant variance in Internet adoption across locations within firms; in general, firms did not adopt basic Internet across all establishments at the same time.

#### 3.2. Model identification

A relationship between adoption of Internet technology and citations between locations does not, in itself, allow us to assert that our findings are consistent with the data-generating processes described above. A general concern with our research design is that there might be unobserved time-varying factors that could be correlated both with Internet adoption and with knowledge flows. For example, a firm-wide effort to increase knowledge flows between researchers could give rise to Internet adoption.

To address these latter concerns, we explore the robustness of our results. First, we explore whether our results appear at the "correct" time. We examine whether a variable capturing dyadic Internet adoption prior to when adoption actually occurs is associated with an increase in citations. If we observe that such a pre-adoption indicator is associated with citation flows, this is evidence that is suggestive of omitted variable bias.

Second, we explore the robustness of our results to the use of instrumental variables. We instrument for Internet adoption using variables that will shift the costs of adopting Internet technology but should have little effect on knowledge flows. Because the instruments are shaped by regional variation, we compute the variables for each of the two locations in the dyad and then take the average. We use two instruments. The first instrument captures local telecommunications costs: It is the year in which the state changed to rate of return (ROR) regulation for telecommunications services. Changes in regulatory policy can influence the likelihood of basic Internet adoption in two ways, in potentially opposite directions. First, by directly lowering the costs of purchasing telecommunications services, they may directly influence the costs of adoption. Second, as Greenstein and Mazzeo (2006) note, this variable can capture local variance in regulatory stringency.

We also instrument using the number of ARPANET nodes in the MSA. The ARPANET was a wide area network that was a predecessor to the Internet. Increases in this variable will represent gains to local familiarity with Internet technologies. (Forman, Goldfarb, and Greenstein (2005, 2008) argue that such local capabilities and expertise can lower the costs of adopting Internet technologies.

As noted earlier, Internet adoption is zero in 1992 and 1994. As a result, we interact each instrument with a dummy variable that is equal to one during 1996 and 1998 and zero otherwise.

Another possibility is that Internet adoption reduces the costs of accessing knowledge from everywhere in the organization, not just locations that are geographically dispersed. In this case, the localization of knowledge flows would not change but the coefficient on β may still be positive because of the firm-wide trend of an increase in citations. To investigate this possibility, we examine the effects of Internet adoption on the likelihood of a citation between inventors within the same MSA and compare it to our baseline estimates.

Another potential issue is that adoption of Internet technology could influence the organizational structure of research operations within the firm. In particular, adoption of IT systems have been shown to influence

the extent to which decision rights within firms are centralized or de-centralized (e.g., McElheran, 2014; Bresnahan et al., 2002). Further, the centralization of research operations within firms has been shown to influence the breadth of search and the extent to which it is tied to the needs of specific business units (Argyres and Silverman, 2004). To address this concern, we rely on the results of prior work that has argued that organizational structure decisions are quasi-fixed in the short run (e.g., Bresnahan et al., 2002) and focus on the short run responses of firms to the initial commercialization of the Internet. In other words, our research design relies on the short run knowledge-sharing strategies of firms in response to the adoption of Internet technology, prior to a point when firms are able to change the organizational structure of their research operations.

### 3.2.1. Potential issues arising from measurement and model specification

Our model estimates may be subject to biases arising from measurement error and model specification. For one, we do not have a direct measure of knowledge flows, and use patent citations to capture knowledge flows. A commonly used assumption is that a citation from patent A to patent B reflects that patent A builds upon the knowledge in patent B (e.g., Jaffe et al., 1993). The advantage of this approach is that organizations are legally obligated to disclose the knowledge that they build upon when they file a patent, and so citations offer a systematic way of tracing knowledge flows within organizations. Further, self-citations have been shown to be positively correlated with firm value (Hall et al., 2005) and so examining the factors that influence self-citations is economically important.

However, there are limitations to using patent citations as a proxy for knowledge flows.[7] In particular, not all citations reflect knowledge flows and some knowledge flows will not be reflected in citations (Griliches, 1990). This problem is exacerbated in research designs that mix self-citations with citations to patents from other firms, which may be driven by different strategic imperatives. Our focus on organizational self-citations will mitigate some of these concerns. The key identifying assumption in our setting is that firms are not changing their self-citation behavior (e.g., as a result of a change in appropriability strategy) simultaneously with Internet adoption.

In our main analysis we focus upon a linear probability model rather than a binary choice model like a probit or logit for several reasons. First, we rely on within variation for identification and removing time invariant heterogeneity in a probit or logit model requires either stronger assumptions (in the case of the probit) or does not enable the computation of marginal effects (in the case of the logit). Further, the computation of marginal effects is less straightforward in nonlinear models, particularly for those including interaction terms (Ai and Norton, 2003), and so our results using the linear probability model are easier to interpret. We view our estimates as a linear approximation to an underlying nonlinear model.

Each firm location pair-year combination appears twice in our data, once with a given firm-location appearing as the citing location and once with it appearing as the cited location. To address standard error concerns related to duplicate values of the same covariates across multiple observations in our sample, we create an index for each firm location pair that is independent of their position (as citing or cited) and cluster our standard errors around that index.

### 3.3. Exploring conditions when the effects of digital technology adoption on knowledge flows will be strongest

We next seek to understand how the effects of Internet adoption will vary based upon the similarity and breadth in research domains

between the source and recipient of the knowledge flows. To do this, we interact our measure of Internet adoption with measures of technological proximity in research areas across locations in the dyad and a measure of the specialization of the research domain of the recipient.

For example, to examine whether the effects of Internet adoption on citation behavior are systematically different for locations working in similar research areas, we estimate the following regression equation:

$$Citation_{ijkt} = \alpha_1 X_{ijkt} + \alpha_2 Z_{ijkt} + \beta Internet_{ijkt} \times Proximity_{ijk} + \mu_{ijk} + \tau_t + \varepsilon_{ijkt} \tag{2}$$

Where $Proximity_{ijkt}$ measures the technological proximity of locations j and k of firm i at time t. Our models for estimating differences in the effects of Internet adoption based on the citing location's specialization in research are estimated similarly.

## 4. Data

Our data come from a variety of sources. We match data on digital technology adoption from a well-known private data source to data on patent citations from the USPTO. We combine these data with information from Compustat (to obtain controls related to R&D and firm size) and from the U.S. Census County Business Patterns data (to obtain data for regional controls).

Our estimation sample is from 1992 to 1998. We believe that this time period is a particularly appealing one to study the effects of digital innovation, because it predates the development and adoption of a profusion of more sophisticated and diverse group of systems and applications such as collaborative tools, instant messaging applications and voice-over-IP communication. The very basic nature of Internet technology at the time implies that (1) little technological and organizational adaptation is required from organizations to benefit from adoption and (2) that the potential underlying mechanisms through which adoption can potentially influence citation patterns are more limited.

### 4.1. Patent data

Within each firm, we use citations between patents invented at different locations as a proxy for within-firm cross-location knowledge spillovers. To do so, we use data on patents filed by multi-establishment US manufacturing firms at the United States Patent and Trademark Office (USPTO). We use the application date as the date for the citing patent because of delays in the application-to-grant period, and because application dates are closer to when the invention occurred (e.g., Griliches, 1990). Our key variable is equal to one where there was a citation from a patent with application date *t* from location *j* to another patent invented in location *k* over the previous ten-year period, and zero otherwise. Our focus on the extensive margin of whether there exist any citations—rather than the count of the number of citations between locations—in part reflects the distribution of our dependent variable: only 8.4% of dyads have a citation between them (6.7% when collaborative patents are excluded, 4% when non-collaborative patents are excluded). This low number is partly explained by a significant share (29%) of firm-locations having no patents in a given year. Conditional on whether the citing location has at least one patent in the focal year, 11.1% of dyads have at least one citation between them (8.9% with collaborative patents excluded, 5.3% with non-collaborative patents excluded). Our citation measure includes examiner citations.[8] While this may overestimate the mean citation propensity overall, it will bias our coefficient estimates only if changes to patent

---

[7] For empirical work exploring the limitations of patent citations as a measure of knowledge flows, see Alcacer and Gittelman (2006) and Roach and Cohen (2013).

[8] For more discussion of this examiner citations and how they influence the interpretation of citations as a measure of knowledge flows, see Alcacer and Gittelman (2006).

citations from examiners are correlated with Internet adoption at the focal firm-location dyads, which looks unlikely.

Our analysis requires us to identify both the firm and location in which a patent is invented.[9] We map patents to firms using the assignee field from the patent and the GVKEY of the Compustat database using the matching files from the National Bureau of Economic Research (NBER) Patent Data Project. Using this procedure, we obtained the universe of patents with a matching GVKEY that were applied for during the period 1990–1998.

As noted above, the unit of analysis in our data will be within-firm location dyad-years, with separate observations based on the citing location in the pair. Rather than use the particular addresses of plants, we assign each plant to an MSA and use the MSA as our unit of location. This reflects a data constraint; the USPTO patent data list only the city and state of an inventor, and so we are unable to identify the particular establishment that an inventor works at within an MSA. Using the city and state of the inventor listed in the patent, we map this information to zip codes and then in turn match zip codes to MSAs. When consolidated MSAs (CMSAs) were present, we used those because they better captured commuting patterns. In regions where inventors resided outside of MSAs, we constructed "phantom MSAs" which consisted of the areas of a state outside of all of the MSAs.

### 4.2. Digital technology data

Our data on Internet adoption comes from a private source, the Harte Hanks Market Intelligence Computer Intelligence Technology database (hereafter, CI database). The database contains a wide range of information related to establishment- and firm-level data on investments in digital technology related to hardware, software, and networking, as well as data related to demographic information such as the number of employees and industry of the establishment and firm. The data have been used in a wide range of studies related to the adoption of IT (e.g., Bresnahan and Greenstein, 1996; Forman et al., 2005) and IT productivity (e.g., Brynjolfsson and Hitt, 2003; Bloom et al., 2012; Nagle, 2018). More closely related to our setting, it has been used to study the role of digital technology investments in reducing the costs associated with economic and geographic isolation in other settings (e.g. Forman, 2005; Forman et al., 2005) and on the effects of digital technology investments on the organization of R&D (Forman and van Zeebroeck, 2012; Forman et al., 2015).

As noted in prior work (e.g., Forman, 2005; Forman et al., 2005), the CI database contains a wide range of information related to an establishment's adoption of digital technology. For this paper, our interest is in exploring the implications of a margin of Internet that lowers communications costs across establishments, but which imposes little direct change on the business processes of organizations.

We consider an establishment to have adopted basic Internet when any one of the following occurs: the establishment reports that it has an Internet Service Provider (ISP); the establishment reports having an internal intranet based on the TCP/IP protocol (Transmission Control Protocol/Internet Protocol); or the establishment reports using the Internet for research purposes. We assume that no establishments have adopted Internet in 1992–1994 as this period predates the launch of the commercial Internet. Among the 37,720 pairs of locations, 0% had adopted basic Internet in 1992 and 1994, 12% had adopted by 1996, and 70% had adopted by 1998.

The CI data are collected at the establishment level. We match establishments to MSAs using the establishment zip code to match our establishment-level IT data to USPTO patent data. Whenever we have several establishments within a given MSA, we consider that a location has adopted as soon as one establishment within the location has done

---

[9] The discussion in this section and the next is a summary of the relevant issues. For further details, see Forman and van Zeebroeck (2012).

---

so. To obtain controls from Compustat such as R&D expenses, we further match the firm identifier in the CI database to a Compustat GVKEY.

### 4.3. Firm-MSA pairs and dependent variable

The focus of our study is to examine the effects of Internet technology adoption on cross-establishment knowledge flows within organizations. We estimate Eq. (1), which allows us to measure whether adoption of basic Internet (in its various forms, either dyadic or citing only) in year $t$ is associated with a citation from location $j$ to location $k$ and vice-versa. To do this, we form the complete set of potential firm-location pairs within an organization, and examine whether there exists a patent in location $j$ (invented during the focal period $t$) that cites a patent invented at location $k$ over the ten-year period preceding time $t$ (and vice-versa). The dataset is symmetric, which means that we keep both combinations of every set of two locations ($j$-$k$ and $k$-$j$). The dependent variable is different in the two configurations: it will reflect citations by patents invented in location $j$ to patents invented in location $k$ in the first case, and citations from $k$-patents to $j$-patents in the second. Our main dependent variable is a binary measure indicating whether establishment $j$ makes at least one citation to a patent invented at establishment $k$ in the preceding 10-year period.

We restrict our estimation sample to firm MSA dyad-year combinations where the firm is in the manufacturing industry (Standard Industrial Classifications 20–40) and to firm-MSAs in which there is at least one patent in two separate years during the period 1992–1998. These conditions are to restrict our sample to establishments engaged in research and patenting activities.

### 4.4. Controls

We control for a variety of firm- and location-specific factors in our regressions. To control for variance in R&D inputs across firms, we compute the flow of R&D spending (in dollars) using Compustat and normalize this figure by dividing total spending by the number of firm-locations in our data. We use the Harte Hanks data to compute firm-location employment as the sum of establishment employment across establishments in the location. Because we do not observe employment in 1992 and 1994 (we do not have CI data for these years), we assume 1996 values for these years. We compute the log of the average employment across the two locations to estimate Eq. (1).

To control for how technological similarity between two establishments influences the likelihood of observing a citation, we compute technological proximity based on Jaffe (1986) and MacGarvie (2006) using the share of patent portfolios that fall in the same technological classes. Specifically, the proximity between locations j and k of firm i is computed as:

$$Prox_{ijkt} = \frac{\sum_{c=1}^{C} P_{ijct} P_{ikct}}{\sqrt{(\sum_{c=1}^{C} P2_{ijct})(\sum_{c=1}^{C} P2_{ikct})}}$$

where C is the total number of technological classes considered, and t is the period over which we compute proximity. Following Benner and Waldfogel (2008), we consider all USPC classes assigned to patents in our sample in order to minimize biases in our measure. Our results are however robust to the use of the main technological class only. We computed proximity over the two-year period (1989–1990) that precedes our analysis period (1991–1998). For some firm-location pairs this variable was undefined because one of the establishments in the pair had no patents during the period considered. In this case we added a dummy variable to indicate that proximity is undefined. In our regressions, rather than the nominal proximity score, we use a dummy that equals to 1 if the proximity score of the focal dyad is above the median in our sample and zero otherwise.

In a similar way, we also compute the degree of technological specialization of each establishment using a Herfindahl index. The

resulting score ranges between 1 (all patents are concentrated in a single class) and 0 (all patents belong to different classes). Here again, we compute the specialization index based on the distribution of the focal establishment's patents across 525 US classes prior to the start of our sample period.[10] If $s_{ijt}$ is the share of patents applied for by establishment $j$ in period $t$ that fall in the $i$th technological class, the Herfindahl index is then given by:

$$H_{jt} = \sum_{i=1}^{n} s_{ijt}^2$$

Both measures, technological proximity among pairs and technological specialization at the establishment level, can be measured at different points in time: simultaneously with the focal observation period (running in-sample measure), over the entire period of analysis (fixed in-sample measure), or over a period preceding our analysis period (pre-sample measure). We have investigated our results using alternative versions of these measures and they are robust.

We further control for other potential sources of heterogeneity. We control for gross innovative output of the dyad and the environment using patent stocks at the dyad and county levels. To account for potential changes in patenting productivity, we compute the dyad's patent stock in the current period (running count) in addition to the stock observed over a moving 10-year period (i.e., cumulative stock of patents invented at any of the two locations over the previous 10 years). We include these measures to ensure that our results are not explained by raw increases in patent output. Additional county-level controls include the share of local employment in manufacturing, local average weekly wages, and the log of local employment.

Table 1 reports descriptive statistics for our main variables.

## 5. Empirical results

Our discussion of results will be organized according to our theoretical framework. To probe the circumstances under which Internet adoption will lead to an increase in knowledge flows, we begin by comparing its impact across the four cases in Fig. 1. We first compare the effects for citations from non-collaborative patents when both locations and the citing location only adopt (cases (A) and (B) in Fig. 1). We next consider the implications of Internet adoption for patents arising from collaborations (cases (C) and (D) in Fig. 1). We explore the robustness of these results to a variety of alternative estimators and robustness checks, including the use of instrumental variables.

We next explore how the effect of Internet on knowledge flows differ based on the overlap in research domains between the source and recipient. In this case, because we are interested in the differences based on proximity and specialization on total citation flows—whether they arise from collaborative or non-collaborative patents—we estimate our models over that sample. Last, we examine whether dyadic v. citing only adoption are differentially able to overcome barriers to knowledge transfer based on proximity and specialization.

### 5.1. Baseline results using non-collaborative patents

We begin by establishing a baseline result between Internet adoption and dyadic citation patterns by estimating regression Eq. (1) using patent citations that do not arise from a collaboration (cases (A) and (B) in Fig. 1). The results are reported in Table 2. Column 1 shows that

**Table 1**
Descriptive statistics.

| Variable | Mean | SD | Min | Max |
|---|---|---|---|---|
| Pair has at least one 10-year citation flow | 0.08 | 0.28 | 0.00 | 1.00 |
| Pair has at least one 10-year citation flow - Collaborative patents excluded | 0.07 | 0.25 | 0.00 | 1.00 |
| Pair has at least one 10-year citation flow - Collaborative patents only | 0.04 | 0.20 | 0.00 | 1.00 |
| Basic Internet in both locations | 0.20 | 0.40 | 0.00 | 1.00 |
| Internet adoption at the citing location in the pair | 0.29 | 0.45 | 0.00 | 1.00 |
| Proximity pre-sample is above median | 0.21 | 0.41 | 0.00 | 1.00 |
| Proximity is not available | 0.45 | 0.50 | 0.00 | 1.00 |
| Specialization (USCL) of the citing location is above median | 0.36 | 0.48 | 0.00 | 1.00 |
| Specialization is not available | 0.26 | 0.44 | 0.00 | 1.00 |
| Log of per-establishment R&D spending | 3.02 | 1.47 | 0.46 | 7.73 |
| Log of patent stock over previous 10 years | 2.84 | 1.64 | 0.00 | 9.02 |
| Log of patent stock in current period | 1.78 | 1.48 | 0.00 | 7.61 |
| Log of establishment employees | 7.69 | 1.13 | 5.30 | 12.04 |
| Share of local employment in manufacturing | 0.20 | 0.06 | 0.04 | 0.52 |
| Local average weekly wages | 543.60 | 85.44 | 306.48 | 848.33 |
| Log of local employment | 13.84 | 0.94 | 10.35 | 15.70 |
| Log of number of local patents | 6.69 | 1.21 | 0.69 | 9.13 |
| Year of change to ROR regulation x after 1996 dummy | 994.78 | 994.79 | 0.00 | 1999.00 |
| Mean number of ARPANET nodes x after 1996 dummy | 0.63 | 1.43 | 0.00 | 9.00 |
| Year of change to Price cap regulation x after 1996 dummy | 47.19 | 47.21 | 0.00 | 99.00 |

*Note*: Number of observations = 37,720.

adoption of Internet technology at both locations is associated with a 1.2 percentage point increase in the likelihood of observing a citation between the two locations (statistically significant at the 5% level). Compared to an average likelihood of observing a citation of 6.72%, this represents a 17.6% increase in the likelihood of observing a citation.[11]

Column 2 shows the results when only the citing location adopts Internet technology. While the result suggests that adoption of Internet technology at only the citing location is associated with a 0.8 percentage point increase in the likelihood of adoption (significant at the 5% level) this may be picking up some of the effects of dyadic Internet adoption. When both dyadic and citing only Internet adoption is included in the same regression (column 3) citing only Internet has no statistically significant effect on the likelihood of a citation while adding adoption at the cited location increases the effect of Internet adoption by 1.0 percentage points (significant at the 10% level). Together, adoption at the citing and cited location is associated with a $1.02 + 0.26 = 1.28$ percentage point in the likelihood of observing a citation (significant at the 1% level). In short, the benefits of Internet adoption on citations from non-collaborative patents arise primarily when the Internet is adopted at both locations.

As often happens in linear probability models (e.g., Athey and Stern, 2002; Agrawal and Goldfarb, 2008), the within $R^2$ value is low across all columns in the table. As a comparison, we include in the table the value of $R^2$ including the explanatory power of the fixed effects, which is higher (around 0.6).

We explore the robustness of our results in Table A3 in the Appendix. Because most of the benefits of non-collaborative patents

---

[10] Our main specification uses 3-digit US classes (525 technological classes), which we believe is a well-balanced level of measurement for specialization at the firm level. We have found, however, consistent results using Hall, Jaffe and Trajtenberg's classification (referred to as "HJT"), which includes 37 sub classes. See van Zeebroeck et al. (2006) for a discussion of the different design choices to be made when computing specialization measures based on patent data and their implications.

[11] Note that we have estimated different versions of Poisson and zero-inflated Poisson models to estimate the impact of Internet adoption on the number (rather than the incidence) of citations. The results systematically showed that most of the implications for Internet adoption in our data occurs at the extensive margin of whether or not a citation occurs, rather than the effects on number of citations. Due to space limitations we do not report these results but they are available from the authors upon request.

**Table 2**
Dyadic v. Browser-based Internet - Non-collaborative cites only.

| | At least one cite where citing is NOT collaborative - Dyadic Internet (1) | At least one cite where citing is NOT collaborative - Only citing-location adoption (2) | At least one cite where citing is NOT collaborative - Both adoption measures (3) | At least one cite where citing is NOT collaborative - Dyadic Internet - IV (4) |
|---|---|---|---|---|
| Basic Internet in both locations | 0.0118** | | 0.0102* | 0.1495* |
| | (0.0047) | | (0.0059) | (0.0783) |
| Internet adoption at the citing location in the pair | | 0.0083** | 0.0026 | |
| | | (0.0041) | (0.0052) | |
| Log of per-establishment R&D spending | 0.0048 | 0.0048 | 0.0048 | 0.0024 |
| | (0.0050) | (0.0050) | (0.0050) | (0.0054) |
| Log of patent stock over previous 10 years | 0.0251*** | 0.0250*** | 0.0251*** | 0.0253*** |
| | (0.0034) | (0.0034) | (0.0034) | (0.0035) |
| Log of patent stock in current period | 0.0222*** | 0.0222*** | 0.0222*** | 0.0209*** |
| | (0.0019) | (0.0019) | (0.0019) | (0.0020) |
| Log of establishment employees | −0.0269*** | −0.0268*** | −0.0269*** | −0.0317*** |
| | (0.0099) | (0.0099) | (0.0099) | (0.0108) |
| Share of local employment in manufacturing | 0.2731 | 0.2684 | 0.2725 | 0.3336 |
| | (0.2220) | (0.2221) | (0.2219) | (0.2351) |
| Local average weekly wages | 0.0005*** | 0.0005*** | 0.0005*** | 0.0005*** |
| | (0.0001) | (0.0001) | (0.0001) | (0.0001) |
| Log of local employment | −0.1458*** | −0.1455*** | −0.1455*** | −0.1319** |
| | (0.0513) | (0.0513) | (0.0513) | (0.0544) |
| Log of number of local patents | 0.0018 | 0.0016 | 0.0016 | −0.0050 |
| | (0.0127) | (0.0128) | (0.0128) | (0.0137) |
| Constant | 1.8325*** | 1.8316*** | 1.8305*** | |
| | (0.6721) | (0.6722) | (0.6721) | |
| $R^2$ | 0.03 | 0.03 | 0.03 | |
| $R^2$ including fixed effects | 0.60 | 0.60 | 0.60 | |
| N | 37,720 | 37,720 | 37,720 | 37,720 |

*Notes*: The dependent variable is the incidence of a patent citation in a non-collaborative patent between the citing firm-MSA and cited firm-MSA in the pair. All regressions include a constant term, firm-pair fixed effects, and time dummies. Robust standard errors, clustered on firm-location pairs, are in parentheses.

  * Significant at the 10% level.
  ** Significant at the 5% level.
  *** Significant at the 1% level.

arise from dyadic Internet adoption, we focus on robustness using that measure. In Table A3 we show that the timing of the effects of dyadic Internet adoption on citations from non-collaborative patents do not appear earlier than they should, are robust to the use of an unbalanced panel, and are robust to different citation windows. One question is whether adoption of Internet technology leads to an increased number of citations in general, not just those across establishments within the same firm. To explore this possibility, in Appendix Table A4 we present results of a regression that explores the effects of Internet adoption on patent citations within the same location at the firm. We find that adoption of Internet technology is not associated with an increase in the likelihood of observing a citation from patents invented in the same MSA.

To further address concerns about omitted variable bias, we present the results of instrumental variables estimates for our baseline dyadic Internet adoption measure in column 4 of Table 2. We estimate two-stage least squares models using the instruments that were described above (year of first change to ROR regulation and number of ARPANET nodes). Table 2 presents the second stage results; we present first stage and second stage results for both instruments together and separately in Appendix Tables A1 and A2.

Column 4 of Table 2 shows that the sign of the second-stage coefficient on basic Internet remains robust to the use of instrumental variables, however there are some differences in both the size and statistical significance of the effects of Internet adoption. The coefficient on dyadic Internet adoption in column 1 is 0.0118 (significant at the 5% level), while the coefficient in the IV regression is 0.1495

(significant at the 10% level). One potential reason for this pattern is that our results may reflect a local average treatment effect. While our results may be valid in the sense that they are uncorrelated with citation patterns but for their impact on basic Internet adoption, it may be that firm location pairs whose Internet adoption behavior is influenced by variance in the instruments will also have a particularly large increase in citations resulting from basic Internet adoption (Angrist et al., 1996).

### 5.2. Baseline results including collaborative patents

Columns 1 through 4 of Table 3 provides mixed evidence that Internet adoption is associated with an increase in the likelihood of citation in patents that arise from collaborations between the citing and cited locations. Perhaps surprisingly, neither dyadic Internet adoption nor Internet adoption at only the citing location is associated with an increase in the likelihood of a citation (columns 1 to 3) in the OLS fixed effect regressions. However, as column 4 indicates, a positive and statistically significant effect does reappear in our instrumental variables estimates, similar in size to that excluding collaborative cites. Interpreting these estimates is difficult: In our setting the likelihood of a citation from a collaboration will increase if either the likelihood of a collaboration increases or the likelihood of a citation conditional on a collaboration increases. Prior work using similar data has demonstrated that the likelihood of cross-location collaborations will increase after the adoption of Internet technology (Forman and van Zeebroeck, 2012). However, we are unable to separately identify these two effects, as

C. Forman, N. van Zeebroeck

**Table 3**
Dyadic v. Browser-based Internet - Collaborative v. All cites.

| | At least one cite where citing is collaborative - Dyadic Internet (1) | At least one cite where citing is collaborative - Only citing-location adoption (2) | At least one cite where citing is collaborative - Both adoption measures (3) | At least one cite where citing is collaborative - Dyadic Internet - IV (4) | Any citation - Dyadic Internet (5) | Any citation - Only citing-location adoption (6) | Any citation - Both adoption measures (7) | Any citation - Dyadic Internet - IV (8) |
|---|---|---|---|---|---|---|---|---|
| Basic Internet in both locations | 0.0045 (0.0041) | | 0.0063 (0.0054) | 0.1471* (0.0667) | 0.0100* (0.0052) | | 0.0097 (0.0064) | 0.1310 (0.0851) |
| Internet adoption at the citing location in the pair | | 0.0007 (0.0035) | −0.0028 (0.0047) | | | 0.0058 (0.0046) | 0.0004 (0.0057) | |
| Log of per-establishment R&D spending | 0.0037 (0.0035) | 0.0038 (0.0035) | 0.0038 (0.0047) | 0.0012 (0.0046) | 0.0021 (0.0055) | 0.0021 (0.0046) | 0.0021 (0.0057) | −0.0000 (0.0059) |
| Log of patent stock over previous 10 years | 0.0233*** (0.0043) | 0.0233*** (0.0043) | 0.0233*** (0.0043) | 0.0235*** (0.0047) | 0.0358*** (0.0055) | 0.0358*** (0.0055) | 0.0358*** (0.0055) | 0.0360*** (0.0059) |
| Log of patent stock in current period | 0.0131*** (0.0026) | 0.0131*** (0.0026) | 0.0131*** (0.0026) | 0.0117*** (0.0028) | 0.0282*** (0.0038) | 0.0282*** (0.0038) | 0.0282*** (0.0038) | 0.0270*** (0.0039) |
| Log of establishment employees | −0.0221** (0.0016) | −0.0220** (0.0016) | −0.0221** (0.0016) | −0.0271*** (0.0017) | −0.0304*** (0.0021) | −0.0303*** (0.0021) | −0.0304*** (0.0021) | −0.0347*** (0.0022) |
| Share of local employment in manufacturing | 0.0846 (0.1687) | 0.0826 (0.1688) | 0.0852 (0.1687) | 0.1473 (0.1872) | 0.3329 (0.2419) | 0.3289 (0.2419) | 0.3329 (0.2418) | 0.3862 (0.2521) |
| Local average weekly wages | 0.0005*** (0.0001) | 0.0005*** (0.0001) | 0.0005*** (0.0001) | 0.0005*** (0.0001) | 0.0007*** (0.0001) | 0.0007*** (0.0001) | 0.0007*** (0.0001) | 0.0007*** (0.0001) |
| Log of local employment | −0.1144** (0.0451) | −0.1147** (0.0451) | −0.1147** (0.0451) | −0.1000** (0.0484) | −0.1853*** (0.0541) | −0.1854*** (0.0541) | −0.1853*** (0.0541) | −0.1732*** (0.0564) |
| Log of number of local patents | −0.0084 (0.0104) | −0.0082 (0.0104) | −0.0082 (0.0104) | −0.0155 (0.0111) | −0.0210 (0.0140) | −0.0211 (0.0141) | −0.0211 (0.0140) | −0.0271* (0.0148) |
| Constant | 1.4914** (0.5981) | 1.4943** (0.5981) | 1.4936** (0.5984) | | 2.4294*** (0.7117) | 2.4301*** (0.7118) | 2.4291*** (0.7118) | |
| $R^2$ | 0.02 | 0.02 | 0.02 | | 0.03 | 0.03 | 0.03 | |
| $R^2$ including fixed effects | 0.58 | 0.58 | 0.58 | | 0.61 | 0.61 | 0.61 | |
| N | 37,720 | 37,720 | 37,720 | 37,720 | 37,720 | 37,720 | 37,720 | 37,720 |

*Notes*: The dependent variable is the incidence of a patent citation between the citing firm-MSA and cited firm-MSA in the pair. Columns 1–4 use the set of patents that arise from a collaboration between the citing and cited locations, while columns 5–8 use the set of all patents. All regressions include a constant term, firm-pair fixed effects, and time dummies. Robust standard errors, clustered on firm-location pairs, are in parentheses.

   \* Significant at the 10% level.
   \*\* Significant at the 5% level.
   \*\*\* Significant at the 1% level.

consistent estimates would require an exclusion restriction that would influence the likelihood of a collaboration but not the likelihood of a citation.

Columns 5 through 8 show the results of Internet adoption on any type of citation, arising either from collaborative or non-collaborative patents. Column 5 shows that for the OLS estimates the statistical and economic significance is somewhere between those for collaborative and non-collaborative patents. In regressions using dyadic Internet, column 5 shows that adopting Internet at both locations is associated with a 1.0 percentage point increase in the likelihood of a citation (significant at 10% level). As in prior tables, adopting Internet only at the citing location has no statistically significant impact on the likelihood of a citation (column 6). Column 7 shows that while the incremental effects of citing only and dyadic Internet are not statistically significant, the overall effect of adopting dyadic Internet are equal to $0.0097 + 0.0004 = 0.0101$ and statistically significant at the 10 percent level. The effect of dyadic Internet is however just below conventional significance levels using our IV estimates (in column 8, the p-value is 0.123).

We have rerun our robustness estimates (as reported in Tables A3 and A4) using any type of citation, arising either from collaborative or non-collaborative patents, as in columns 5 to 8. The results are qualitatively similar, and therefore also support a causal interpretation of the relationship between dyadic Internet adoption and citation flows between establishments. To eliminate potential concerns of how our estimates might be influenced by self-selection into collaborative or non-collaborative patents, in all future tables we will use the estimates for this version of the dependent variable, using citations arising from any type of patent, either collaborative or non-collaborative.

### 5.3. Exploring the effects of technological proximity and specialization

In this section we study whether the effects of adopting basic Internet on citation patterns are greater when the citing and cited locations are technologically proximate and when the citing location is technologically specialized. Our identification assumptions are somewhat weaker in this section than earlier. Here our primary assumption is that there do not exist unobserved variables that are correlated with Internet and citations and that are differentially trending for pairs where there is high proximity and where knowledge is specialized. Table 4 shows the results of estimating model 2 and its analog for the proximity and specialization tests. As noted above, for these models we use citations arising from any type of patent (collaborative or non-collaborative) and in this table examine the effects of dyadic Internet adoption only.

Column 2 shows heterogeneity in the effects of Internet based on the proximity of research areas between the locations in the pair. The results show that the effects of Internet are much stronger for establishments who are in similar fields. Our proximity measure requires that both locations in the pair patent at least once in the pre-sample (1989–1990) period over which proximity is calculated. For some pairs of locations there were no patents over this period with which to compute the proximity measure: We were unable to compute a proximity measure for 45.1% of observations in our sample. To control for the presence of location pairs for which we are unable to compute a proximity measure, we include an addition interaction between Internet and an indicator that proximity is not available.

The point estimate for our variable capturing Internet x Proximity pre-sample is above median is 0.0497 (significant at the 1% level), implying that establishment pairs who adopt Internet and who are technologically proximate see a 5.10 percentage point ( $= 0.0013 + 0.0497$ ) increase in the likelihood of a citation after adopting Internet technology. In contrast, pairs who adopt Internet who are not technologically proximate see no increase in the likelihood of observing a citation.

We have probed the robustness of this result in several ways. First,

we have computed our proximity measure using in-sample data. We have both allowed proximity to vary in-sample and also created a time-invariant version of the variable computed over our entire period of analysis (1991–1998). Our results are robust to both approaches. Also, one alternative explanation for our results is that technological proximity between two locations is proxying for prior collaborations between them. That is, locations that are technologically proximate will have also collaborated before, and such locations will benefit disproportionately from Internet adoption because prior ties between the locations will increase the likelihood of a citation, other things equal (Singh, 2005). To address this possibility, in Appendix Table A5 (in the first column) we present results from regressions that include proximity and presence of prior collaborations in the same regression. We find that when the interactions of Internet and both proximity and prior collaborations are included in the same regression, the statistical and economic significance of the Internet and proximity interaction remains similar to that as in Table 5 while there is no effect of Internet on citation flows for pairs of locations that have collaborated before.

Column 3 provides the results of our proximity interactions using instrumental variables. To construct the instruments, we use the same two instruments as in column 4 of Table 2 and interact them with our proximity measures. Thus, we have six instruments in total: the two instruments in Table 2 along with their interactions with proximity and proximity not available. The results are directionally similar to those without instruments, although the coefficients of adopting dyadic Internet and its interaction with proximity are larger in magnitude than those in column 2.

In columns 4–5 we show the results of regressions that allow the marginal effect of Internet to vary based on the specialization of research interests in the establishment. As described above, our baseline measure of specialization is based upon 525 US classes. As was the case for our proximity calculations, there are some citing locations for which there were no patents in the pre-sample period with which to calculate a measure of specialization. As a result, we construct an additional term which is equal to one when we are unable to compute a specialization measure.[12] This happens for 25% of our sample. We interact this with our Internet variable to control for differences in the effects of Internet adoption among pairs for whom we are unable to compute specialization.

Column 4 shows that the effects of Internet adoption decrease significantly when the citing location is highly specialized in its research. When the citing establishment is not specialized, pairs of locations who adopt Internet see a statistically significant (at the 1% level) 3.03 percentage point increase in the likelihood of a citation. When the citing establishment is specialized, there is no statistically significant effect of Internet on the likelihood of observing a citation.

We have explored the robustness of this result to different ways of measuring specialization. Specifically, we explore robustness to computing specialization using the more aggregated Hall-Jaffe-Trajtenberg (Hall et al., 2001) technological areas. Our results are robust to this change. Column 5 of Table 4 also shows that our results are qualitatively similar when we use instrumental variables.[13]

Finally, columns 6 and 7 report additional estimates that include proximity and specialization in the same regression. Column 6 shows that the results of proximity and specialization remain statistically and economically significant when included in the same regression, however the economic impact of specialization weakens. We again run

---

[12] As noted earlier, our sample only include firm-locations that patent during our sample period. However, this is a within-sample condition. As a result, a firm-location could be included in our sample and still not patent during the pre-sample period and therefore lack a measure of specialization.

[13] As was the case for proximity, we use our two instruments in Table 2 and interact those with our measures of specialization and specialization not available.

**Table 4**
Proximity and Specialization (all cites/dyadic Internet).

| | Baseline (1) | With proximity pre-sample (2) | With proximity pre-sample - IV (3) | With specialization pre-sample (4) | With specialization pre-sample - IV (5) | With proximity & specialization pre-sample (6) | With proximity & specialization pre-sample - IV (7) |
|---|---|---|---|---|---|---|---|
| Basic Internet in both locations | 0.0100* (0.0052) | 0.0013 (0.0060) | 0.0677 (0.0828) | 0.0303*** (0.0071) | 0.1566* (0.0839) | 0.0119 (0.0074) | 0.0901 (0.0850) |
| Internet X Proximity pre-sample is above median | | 0.0497*** (0.0114) | 0.1156*** (0.0208) | | | 0.0445*** (0.0115) | 0.1031*** (0.0211) |
| Internet x Proximity not available | | −0.0058 (0.0065) | −0.0139 (0.0139) | | | 0.0038 (0.0084) | −0.0057 (0.0159) |
| Internet X Specialization (USCL) above median | | | | −0.0275*** (0.0074) | −0.0692*** (0.0135) | −0.0181** (0.0073) | −0.0459*** (0.0137) |
| Internet X Specialization not applicable | | | | −0.0428*** (0.0079) | −0.0876*** (0.0158) | −0.0286*** (0.0095) | −0.0455*** (0.0165) |
| Log of per-establishment R&D spending | 0.0021 (0.0055) | 0.0016 (0.0055) | −0.0005 (0.0058) | 0.0022 (0.0055) | 0.0005 (0.0058) | 0.0017 (0.0055) | −0.0002 (0.0058) |
| Log of patent stock over previous 10 years | 0.0358*** (0.0038) | 0.0386*** (0.0039) | 0.0426*** (0.0041) | 0.0368*** (0.0038) | 0.0379*** (0.0038) | 0.0380*** (0.0039) | 0.0420*** (0.0041) |
| Log of patent stock in current period | 0.0282*** (0.0021) | 0.0288*** (0.0021) | 0.0290*** (0.0023) | 0.0297*** (0.0021) | 0.0303*** (0.0024) | 0.0295*** (0.0021) | 0.0302*** (0.0024) |
| Log of establishment employees | −0.0304*** (0.0109) | −0.0266** (0.0109) | −0.0242** (0.0117) | −0.0289*** (0.0108) | −0.0304*** (0.0114) | −0.0262** (0.0109) | −0.0230* (0.0116) |
| Share of local employment in manufacturing | 0.3329 (0.2419) | 0.3240 (0.2414) | 0.3461 (0.2486) | 0.2998 (0.2407) | 0.3035 (0.2473) | 0.3093 (0.2409) | 0.3084 (0.2467) |
| Local average weekly wages | 0.0007 (0.0001) | 0.0006*** (0.0001) | 0.0005*** (0.0001) | 0.0006*** (0.0001) | 0.0006*** (0.0001) | 0.0006*** (0.0001) | 0.0005*** (0.0001) |
| Log of local employment | −0.1853*** (0.0541) | −0.1775*** (0.0537) | −0.1591*** (0.0545) | −0.1760*** (0.0537) | −0.1574*** (0.0551) | −0.1748*** (0.0536) | −0.1555*** (0.0542) |
| Log of number of local patents | −0.0210 (0.0140) | −0.0200 (0.0140) | −0.0225 (0.0145) | −0.0204 (0.0140) | −0.0244* (0.0146) | −0.0196 (0.0140) | −0.0213 (0.0144) |
| Constant | 2.4294*** (0.7117) | 2.3189*** (0.7056) | | 2.3109*** (0.7070) | | 2.2889*** (0.7042) | |
| $R^2$ | 0.03 | 0.03 | | 0.03 | | 0.04 | |
| $R^2$ including fixed effects | 0.61 | 0.61 | 0.61 | 0.61 | 0.61 | 0.61 | 0.61 |
| N | 37,720 | 37,720 | 37,720 | 37,720 | 37,720 | 37,720 | 37,720 |

*Notes*: The dependent variable is the incidence of a patent citation in any patent (collaborative or non-collaborative) between the citing firm-MSA and cited firm-MSA in the pair. All regressions include a constant term, firm-pair fixed effects, and time dummies. Robust standard errors, clustered on firm-location pairs, are in parentheses.

\* Significant at the 10% level.
\*\* Significant at the 5% level.
\*\*\* Significant at the 1% level.

**Table 5**
Proximity and Specialization: Dyadic v. Browser-based (all cites).

| | Proximity - Both adoption measures (1) | Specialization - Both adoption measures (2) | Proximity & specialization - Both adoption measures (3) |
|---|---|---|---|
| Basic Internet in both locations | 0.0016 | 0.0223** | 0.0054 |
| | (0.0084) | (0.0108) | (0.0118) |
| Internet adoption at the citing location in the pair | −0.0000 | 0.0089 | 0.0073 |
| | (0.0073) | (0.0094) | (0.0102) |
| Internet X Proximity pre-sample above median | 0.0387** | | 0.0369** |
| | (0.0181) | | (0.0184) |
| Internet x Proximity not available | −0.0017 | | 0.0091 |
| | (0.0111) | | (0.0143) |
| Internet at citing location X Proximity above median | 0.0121 | | 0.0085 |
| | (0.0158) | | (0.0161) |
| Internet at citing location X Proximity not available | −0.0046 | | −0.0059 |
| | (0.0099) | | (0.0127) |
| Internet X Specialization (USCL) above median | | −0.0142 | −0.0065 |
| | | (0.0127) | (0.0128) |
| Internet X Specialization not applicable | | −0.0307** | −0.0233 |
| | | (0.0143) | (0.0170) |
| Internet at citing location X Specialization above median | | −0.0144 | −0.0126 |
| | | (0.0114) | (0.0114) |
| Internet at citing location X Specialization not applicable | | −0.0132 | −0.0058 |
| | | (0.0126) | (0.0148) |
| Log of patent stock over previous 10 years | 0.0391*** | 0.0369*** | 0.0385*** |
| | (0.0039) | (0.0038) | (0.0039) |
| Log of per-establishment R&D spending | 0.0015 | 0.0022 | 0.0016 |
| | (0.0055) | (0.0055) | (0.0055) |
| Log of patent stock in current period | 0.0290*** | 0.0299*** | 0.0297*** |
| | (0.0021) | (0.0021) | (0.0022) |
| Log of establishment employees | −0.0264** | −0.0287*** | −0.0260** |
| | (0.0109) | (0.0108) | (0.0109) |
| Share of local employment in manufacturing | 0.3198 | 0.2932 | 0.3009 |
| | (0.2412) | (0.2402) | (0.2405) |
| Local average weekly wages | 0.0006*** | 0.0006*** | 0.0006*** |
| | (0.0001) | (0.0001) | (0.0001) |
| Log of local employment | −0.1751*** | −0.1738*** | −0.1717*** |
| | (0.0536) | (0.0537) | (0.0535) |
| Log of number of local patents | −0.0201 | −0.0207 | −0.0198 |
| | (0.0140) | (0.0140) | (0.0140) |
| Constant | 2.2889*** | 2.2851*** | 2.2508*** |
| | (0.7047) | (0.7064) | (0.7032) |
| $R^2$ | 0.03 | 0.03 | 0.04 |
| $R^2$ including fixed effects | 0.61 | 0.61 | 0.61 |
| N | 37,720 | 37,720 | 37,720 |

*Notes*: The dependent variable is the incidence of a patent citation in any patent (collaborative or non-collaborative) between the citing firm-MSA and cited firm-MSA in the pair. All regressions include a constant term, firm-pair fixed effects, and time dummies. Robust standard errors, clustered on firm-location pairs, are in parentheses.
*Significant at the 10% level.
** Significant at the 5% level.
*** Significant at the 1% level.

instrumental variable regressions in column 7; comparing columns 6 and 7 shows that the economic significance of Internet adoption and its interaction with proximity and specialization are larger (in absolute) value when instrumental variables are used.

As noted above, our approach includes pairs of firm-locations for which proximity cannot be computed. These pairs may be systematically different in unobservable ways from pairs for which we are able to compute proximity and specialization. In particular, locations in pairs for which we are unable to compute these measures may be less R&D-intensive or productive. Dropping these pairs may engender a selection bias in our estimates; therefore, our baseline approach is to include a control for these cases rather than dropping them. However, we also re-estimated the regression model excluding them. The results of these regressions are reported in table A5 in appendix. The results for proximity are quite robust to this change as are the results for specialization that do not include proximity in the regression. However, we do find that the (negative) marginal effect of Internet adoption for pairs

that exhibit specialization is weaker in the regressions that also include a measure for proximity.

*5.4. Exploring the effects of dyadic and browser-based search*

Table 5 presents results where we allow the marginal effects of Internet to differ based on variation in proximity and specialization and whether or not one (citing only) or both (dyadic) locations adopt. We do this for two reasons. First, columns 6 and 7 of Table 3 suggests that citing only Internet has little effect on citation flows from all patents on average; we wish to see whether there is heterogeneity in this effect based on proximity and specialization (i.e., whether citing Internet only has an effect when there is similarity in the knowledge bases between the citing and cited locations). Second, by decomposing the effects of Internet in this way, we can see whether, by facilitating electronically enabled interpersonal interactions, adoption at the cited location decreases the importance of proximity and specialization.

Column 1 of Table 5 shows that the incremental effects of Internet adoption at the citing (-0.0000) or both (0.0016) locations are not statistically significant at conventional levels in the absence of technological proximity. Similarly, the effects of Internet adoption when proximity is present are also not statistically significant when only the citing location adopts. In contrast, the incremental effects of adopting in both locations are stronger in the presence of proximity; the coefficient estimate for the interaction of dyadic Internet adoption with high proximity is 0.0387 is statistically significant at the 5% level. However, to identify the effects of dyadic Internet adoption in the presence of proximity we must take linear combinations. Overall, when both locations adopt and there is technological proximity between establishments, the likelihood of a citation increases by 5.08 ( = 0.0387 + 0.0121) percentage points (significant at the 1% level). In short, these results show that adoption of Internet technology at both locations will have an effect on citation behavior only when the two locations are technologically proximate. Adoption of Internet technology at the citing location only does not have an effect on behavior.

Column 2 shows differences in results based on the extent to which the citing location is technologically specialized and whether one or both locations adopt. Once again, adoption at the citing location only is not associated with an increase in the likelihood of a citation between the locations, even when the citing location is not specialized. When the cited location also adopts and the citing location is not specialized, this leads to an additional 2.23 percentage point increase in the likelihood of a citation after Internet adoption (significant at the 5% level). In sum, when both the citing and cited locations adopt Internet and the citing location is not specialized, this leads to a 3.12 ( = 0.0223 + 0.0089) percentage point increase in the likelihood of a citation (significant at the 1% level). In contrast, when the citing location is specialized, neither adoption at the citing or cited location (or both) influences citation behavior.

Column 3 provides results when the effects of both proximity and specialization are included in the same regression. Once again, adoption of citing Internet only has no effect on behavior, even when specialization is low and there are overlapping knowledge bases across the two locations. Further, even when both locations adopt, Internet adoption only leads to an increase in citation flows when the two areas share similar research domains. The incremental effect of adopting at the cited location in addition to the citing when specialization is low and proximity is high leads to a 4.23 percentage point increase in the likelihood of a citation ( = 0.0054 + 0.0369, significant at the 1% level). The total effect of adopting Internet at both locations when proximity is high and specialization is low is equal to a 5.81 percentage point increase ( = 0.0054 + 0.0369 + 0.0073 + 0.0085, significant at the 1% level). When both proximity and specialization are both high, the overall effect of adopting Internet at both locations is equal to a 3.9 percentage point increase ( = 0.0054 + 0.0369 – 0.0065 + 0.0073 + 0.0085 – 0.0126, significant at the 1% level).

We attempted to rerun the models above using our instrumental variables approach. However, given the large number of potentially endogenous variables in these models, we were unable to capture the independent effects of citing and cited Internet adoption on behavior.

Overall, the results show that Internet adoption leads to an increase in the likelihood of a citation flow only when there is adoption at both locations and only when there is proximity between the two locations. Specialization also influences behavior, but seems to have a weaker impact when the effects of proximity and specialization are included together in the same regression. We do not find evidence that adoption of Internet at the citing location only influences citation behavior, even when proximity is high and specialization is low.

The results above examine the effects of Internet adoption on citations arising from any type of patent. We re-estimated the models for patents arising only from patents that do not involve a collaboration between the two locations. While this approach does involve selecting upon the type of patent, it also allows us to isolate more directly how technology adoption influences search and transfer behavior. The results are reported in table A6. They are in many ways qualitatively similar to those reported in Table 5, with one exception. The results in table A6 show that citing only Internet adoption leads to a statistically significant increase in the likelihood of a citation when proximity is high and specialization is low, albeit an increase that is smaller than for dyadic Internet adoption. While we are unable to isolate the reason for this difference, one reason may be that citing only Internet may decrease the costs of certain types of knowledge flows but, as shown by Forman and van Zeebroeck (2012), will not increase the likelihood of a collaboration.

## 6. Discussion and implications

By analyzing the impact of Internet adoption on R&D knowledge flows within multi-establishment manufacturing firms in the US, this paper investigates how digital technology influences the cross-location transfer of knowledge within firms. Our results show that adoption of an early form of Internet technology significantly increases the likelihood of observing a patent citation between geographically dispersed R&D locations within the same firm, however there is considerable variance when and where this occurs. First, our results show that the economic and statistical effects of adopting dyadic Internet are much stronger than those of adopting citing Internet only. Lower costs of accessing static documents do not have as significant an impact as dyadic adoption that facilitates communication across locations. Second, our results show that these effects are significantly stronger when the citing and cited locations share similar knowledge bases and when the citing location is not specialized in its research.

In sum, these findings are consistent with those of earlier work that has suggested that there are considerable barriers to the transfer and absorption of knowledge across geographic locations. Our work is unique in its ability to separately capture the effects of different types of technology and different modalities of transfer (collaborative v. noncollaborative patents), as well as the effect of proximity and specialization of knowledge bases. Digital technology helps to lower some of the barriers to knowledge transfer, but transfer of knowledge in our setting requires the sort of person-to-person interactions that are facilitated by dyadic technology adoption. Further, even these person-to-person interactions are insufficient to overcome the barriers to transfer and absorption of knowledge across locations with dissimilar knowledge bases. This clearly points at the difficulty of transferring tacit knowledge (requiring tacit to explicit conversions), especially across fields (since they require common vocabulary and norms).

### 6.1. Research implications

Our results show that Internet adoption will increase knowledge flows between locations only when they share a common knowledge base. These results have implications for our understanding of the impact of the digitization of innovation on the rate and direction of inventive activity. In particular, the extent of an economy's ability to recombine new ideas will have important implications for economic growth (Weitzman, 1998). It is widely believed that the digitization of innovation can play an important role in this recombination (Romer, 2008; Brynjolfsson and McAfee, 2014). However, our results suggest that early forms of Internet technology reinforced existing connections among vertical groups of technologies, rather than facilitating recombination of ideas across technology areas.

Earlier work has suggested that digital technology could have implications for the distance between different individuals or different groups with different ex ante types. Namely, because the benefits and costs of a communication link between individuals can be influenced by distance, individuals can endogenously sort into similar groups, which may have implications for collaborations and recombination in science (Van Alsytne and Brynjolfsson, 1996, 2005; Rosenblat and Mobius,

2004). To our knowledge, we provide the first direct evidence of the implications of digital technology adoption on the similarity of the knowledge flows that a scientist draws from. It is important to recognize, however, our distinctions from this prior work. While prior models emphasize the effects of digital technology on the formation of connections at the individual level when scientists have flexibility in determining partnerships, our findings will be shaped by the organization of the research operation within the firm, in particular the extent to which it is centralized or decentralized (Argyres and Silverman, 2004). They show that adoption of digital technology may reinforce existing competencies even in environments where scientists may have less flexibility with regard to their collaborators.

Newer Internet-based technologies may loosen some of the constraints that shape our findings. For example, recent patent search tools allow search approaches that are more flexible than that based on keywords and patent classes (Martin, 2016; Huc, 2017). This may make it easier to find knowledge from related, but different, technological areas. More recent technologies such as online forums may be more effective at facilitating one-to-many communication patterns and allow knowledge searchers to access information from users with whom they may not have existing social connections (McAfee, 2009). Broadcast search such as that enabled by online competitions such as those enabled by InnoCentive may enable seekers of knowledge to access diverse knowledge sources (Jeppesen and Lakhani, 2010). On the other hand, newer applications may reinforce the patterns observed in our data. For example, the same filtering technologies that allow more flexible searches may encourage researchers to more narrowly define searches to areas of related expertise (Van Alstyne and Brynjolfsson, 2005). We also note that the tools we examine in this study continue to be widely used today: In particular, email remains a widespread collaboration tool (Roe, 2018) and is likely to continue to shape knowledge sharing for some time in the future.

In short, this paper has identified a fact: that the effects of digital technology on increasing geographically dispersed knowledge flows are greater between inventors who are working in related fields. It has identified a range of potential reasons for this fact, ranging from factors that inhibit knowledge search and transfer across more distant fields to those that may inhibit absorption. These findings suggest important limitations to the ability of digital technology to increase knowledge flows. It remains however to be seen whether later generations of digital technology will inhibit or reinforce these barriers. We leave this as a question for future research.

Our research examines the short run change in knowledge flows resulting from the adoption of digital technology, holding the structure of research operations as fixed. However, the knowledge flows within firms will be influenced by the organization of its research operations, including the extent to which these operations are centralized or decentralized (Argyres and Silverman, 2004) and the strength of internal linkages across locations (Alcacer and Zhao, 2012). Our research suggests that adoption of IT systems will encourage knowledge flows between scientists working in similar areas. These tendencies may be reinforced in an environment of decentralized research operations that emphasize local search that is tied to existing capabilities (Argyres and Silverman, 2004). More work is needed in this area, but our results complement recent work that has provided evidence of three-way complementarities between data analytics, decentralization of research operations, and research operations that emphasize recombination of existing technologies (Wu et al., 2018).

### 6.2. Limitations

Our analysis is subject to limitations. For one, as noted above, we study a specific set of Internet technologies over a specific time period. The impact of the adoption of Internet technologies on knowledge flows may be different over a later time period. Second, while the nature of our data are unique in that they enable us to study the implications of a change in communication costs on knowledge flows over a large and diverse sample, our unit of analysis makes it difficult for us to pin down exactly the mechanism through which Internet adoption changes behavior.

A related issue is identification of model parameters. We have pursued a variety of analyses to provide a causal interpretation of our results. However, we are also limited by the nature of our data in the types of analyses that we can perform. We look forward to future research that investigations these questions in other settings using other data.

A last limitation is that our research focuses primarily on intra-firm knowledge flows. Future work should examine how the adoption of digital technology influences knowledge flows between institutions such as firms, universities, or research institutes. Research on the latter entities would provide new insights into the 'balkanization of science' hypothesis. As noted above, our focus on firms means that scientists and inventors must focus their research efforts on areas that contribute to the broader product strategy of the firm. Thus, while our results can certainly be influenced by barriers to search, transfer, and absorption costs across distant fields, they are less likely to be influenced by purposeful changes in research interests and connections enabled by technology. However, researchers in universities, for example, will have academic freedom on where to focus their research efforts, and so the risks that digital technology adoption will facilitate changes in search strategies based on preferences will be significant (Rosenblat and Mobius, 2004). We hope our findings encourage future research in this important area.

### Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.respol.2018.10.021.

### References

Agrawal, A., Goldfarb, A., 2008. Restructuring research: communication costs and the democratization of university innovation. Am. Econ. Rev. 98 (4), 1578–1590.

Ai, C., Norton, E.C., 2003. Interaction terms in logit and probit models. Econ. Lett. 80 (1), 123–129.

Alavi, M., Leidner, D., 2001. Knowledge management and knowledge management systems: conceptual foundations and research issues. Mis Q. 25 (1), 107–136.

Alcacer, J., Gittelman, M., 2006. Patent citations as a measure of knowledge flows: the influence of examiner citations. Rev. Econ. Stat. 88 (4), 774–779.

Alcacer, J., Zhao, M., 2012. Local R&D strategies and multilocation firms: the role of internal linkages. Manage. Sci. 58 (4), 734–753.

Allen, T.J., 1977. Managing the Flow of Technology. MIT Press, Cambridge.

Anderson, E.G., Chandrasekaran, A., Davis-Blake, A., Parker, G.G., 2018. Managing distributed product development projects: integration strategies for time-zone and language barriers. Inf. Syst. Res. 29 (1), 42–69.

Angrist, J.D., Imbens, G.W., Rubin, D.B., 1996. Identification of causal effects using instrumental variables. J. Am. Stat. Assoc. 91, 444–455.

Argyres, N.S., Silverman, B.S., 2004. R&D, organizational structure, and the development of corporate technological knowledge. Strateg. Manage. J. 25 (8-9), 929–958.

Athey, S., Stern, S., 2002. The impact of information technology on emergency health care outcomes. RAND J. Econ. 33 (3), 399–432.

Benner, M., Waldfogel, J., 2008. Close to you? Bias and precision in patent-based measures of technological proximity. Res. Policy 37 (9), 1556–1567.

Bercovitz, J., Feldman, M., 2011. The mechanisms of collaboration in inventive teams: composition, social networks, and geography. Res. Policy 40, 81–93.

Bloom, N., Sadun, R., Van Reenen, J., 2012. Americans do IT better: US multinationals and the productivity miracle. Am. Econ. Rev. 102 (1), 167–201.

Bresnahan, T., Brynjolfsson, E., Hitt, L.M., 2002. Information technology, workplace organization, and the demand for skilled labor: firm-level evidence. Q. J. Econ. 117 (1), 339–376.

Bresnahan, T., Greenstein, S., 1996. Technical progress and co-invention in computing and in the uses of computers. Brook. Pap. Econ. Act. Microecon. 1996, 1–83.

Brynjolfsson, E., Hitt, L., 2003. Computing productivity: firm-level evidence. Rev. Econ. Stat. 85 (4), 793–808.

Brynjolfsson, E., McAfee, A., 2014. The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Machines. MIT Press, Cambridge.

Catalini, C., 2018. Microgeography and the direction of inventive activity. Forthcoming. Manage. Sci.

Cohen, W., Levinthal, D., 1989. Innovation and learning: the two faces of R&D. Econ. J. 99 (397), 569–596.

Cohen, W., Levinthal, D., 1990. Absorptive capacity: a new perspective on learning and

innovation. Adm. Sci. Q. 35, 128–152.

Ding, W., Levin, S., Stephan, P., Winkler, A., 2010. The impact of information technology on academic scientists' productivity and collaboration patterns. Manage. Sci. 56 (9), 1439–1461.

Feldman, M.P., Kogler, D.F., 2010. Stylized facts in the geography of innovation. In: In: Hall, B.H., Rosenberg, N. (Eds.), Handbook of the Economics of Innovation 1. Elsevier, Amsterdam, pp. 381–410.

Fleming, L., 2001. Recombinant uncertainty in technological search. Manage. Sci. 47 (1), 117–132.

Forman, C., 2005. The corporate digital divide: determinants of Internet adoption. Manage. Sci. 51 (4), 641–654.

Forman, C., Goldfarb, A., Greenstein, S., 2005. How did location affect the adoption of the commercial internet? Global village vs. urban density. J. Urban Econ. 58 (3), 389–420.

Forman, C., Goldfarb, A., Greenstein, S., 2008. Understanding the inputs into innovation: do cities substitute for internet firm resources? J. Econ. Manag. Strategy 17 (2), 295–316.

Forman, C., Goldfarb, A., Greenstein, S., 2015. Information technology and the distribution of inventive activity. In: Jaffe, Adam, Jones, Ben (Eds.), The Changing Frontier: Rethinking Science and Innovation Policy. University of Chicago Press.

Forman, C., van Zeebroeck, N., 2012. From wires to partners: how the Internet has fostered R&D collaborations within firms. Manage. Sci. 58 (8), 1549–1568.

Galbraith, J.R., 1973. Designing Complex Organizations. Addison-Wesley, Reading, MA.

Garicano, L., 2000. Hierarchies and the organization of knowledge in production. J. Polit. Econ. 108 (5), 874–904.

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

Grant, R.M., 1996. Toward a knowledge-based theory of the firm. Strateg. Manage. J. 17 (S2), 109–122.

Gray, J., Siemsen, E., Vasudeva, G., 2015. Colocation still matters: conformance quality and the interdependence of R&D and manufacturing in the pharmaceutical industry. Manage. Sci. 61 (11), 2760–2781.

Greenstein, S., Mazzeo, M., 2006. The role of differentiation strategy in local telecommunication entry and market evolution: 1999-2002. J. Ind. Econ. 54 (3), 323–350.

Griliches, Z., 1990. Patent statistics as economic indicators: a survey. J. Econ. Lit. 28 (4), 1661–1707.

Gupta, A., Govindarajan, V., 2000. Knowledge flows within multinational corporations. Strateg. Manage. J. 21, 473–496.

Hall, B., Jaffe, A., Trajtenberg, M., 2001. The NBER Patent Citations Data File: Lessons, Insights, and Methodological Tools. NBER Working Paper 8498.

Hall, B., Jaffe, A., Trajtenberg, M., 2005. Market value and patent citations. RAND J. Econ. 36 (1), 16–38.

Hall, B., Lotti, F., Mairesse, J., 2013. Evidence on the impact of R&D and ICT investments on innovation and productivity in Italian firms. Econ. Innov. New Technol. 22 (3).

Hansen, M.T., 1999. The search-transfer problem: the role of weak ties in sharing knowledge across organizational subunits. Adm. Sci. Q. 44 (1), 82–111.

Henderson, R., Clark, K.B., 1990. Architectural innovation: the reconfiguration of existing product technologies and the failure of established firms. Adm. Sci. Q. 35 (1), 9–30.

Henderson, R., Cockburn, I., 1994. Measuring competence? Exploring firm effects in pharmaceutical research. Strateg. Manage. J. 15, 63–84.

Huc, Olivier, 2017. The Challenges of Patent Knowledge Management – Part 1. Blog Post, Available at https://www.linkedin.com/pulse/challenges-patent-knowledge-management-part-1-olivier-huc/.

Jaffe, A., Trajtenberg, M., Henderson, R., 1993. Geographic localization of knowledge spillovers as evidenced by patent citations. Q. J. Econ. 108 (3), 577–598.

Jeppesen, L.B., Lakhani, K.R., 2010. Marginality and problem-solving effectiveness in broadcast search. Organ. Sci. 21 (5), 1016–1033.

Jones, B.F., 2009. The burden of knowledge and the "Death of the renaissance man": is innovation getting harder? Rev. Econ. Stud. 76 (1), 283–317.

Jordan, J., Jones, P., 1997. Assessing your company's knowledge management style. Long Range Plann. 30 (3), 392–398.

Kleis, L., Chwelos, P., Ramirez, R., Cockburn, I., 2012. Information technology and intangible output: the impact of IT investment on innovation productivity. Inf. Syst. Res. 23 (1), 42–59.

Kogut, B., Zander, U., 1992. Knowledge of the firm, combinative capabilities, and the replication of technology. Organ. Sci. 3 (3), 383–397.

Leonard-Barton, D., 1988. Implementation as mutual adaptation of technology and organization. Res. Policy 17 (5), 251–267.

Leiponen, A., Helfat, C., 2011. Location, decentralization, and knowledge sources for innovation. Organ. Sci. 22 (3), 641–658.

Levina, N., Vaast, E., 2005. The emergence of boundary spanning competence in practice: implications for implementation and use of information systems. MIS Q. 29 (2), 335–363.

MacGarvie, M., 2006. Do firms learn from international trade? Rev. Econ. Stat. 88 (1), 46–60.

Martin, J., 2016. The Third Age of Patent Search Tools. Blog Post. Available at https://

www.innography.com/blog/post/the-third-age-of-patent-search-tools.

McAfee, A., 2009. Enterprise 2.0: How to Manage Social Technologies to Transform Your Organization. Harvard Business School Press, Boston.

McElheran, K., 2014. Delegation in multi-establishment firms: evidence from I.T. Purchasing. J. Econ. Manag. Strategy 23 (2), 225–258.

Miller, D.J., Fern, M.J., Cardinal, L.B., 2007. The use of knowledge for technological innovation within diversified firms. Acad. Manag. J. 50 (2), 308–326.

Moenaert, R.K., Souder, W.E., 1990. An information transfer model for integrating marketing and R&D personnel in new product development projects. J. Prod. Innov. Manage. 7 (2), 91–107.

Nagle, F., 2018. Open source software and firm productivity. Forthcoming. Manage. Sci.

Nambisan, S., Lyytinen, K., Majchrzak, A., Song, M., 2017. Digital innovation management: reinventing innovation management research in a digital world. MIS Q. 41 (1), 223–238.

Nelson, R., Winter, S., 1982. An Evolutionary Theory of Economic Change. Harvard University Press, Cambridge, MA.

Nonaka, I., 1991. The knowledge-creating company. Harv. Bus. Rev. 69 (6), 96–104.

Nonaka, I., Takeuchi, H., 1995. The Knowledge Creation Company: How Japanese Companies Create the Dynamics of Innovation. Oxford University Press.

Offsey, S., 1997. Knowledge management: linking people to knowledge for bottom line results. J. Knowl. Manag. 1 (2), 113–122.

Penner-Hahn, J., Shaver, J.M., 2005. Does international research and development increase patent output? An analysis of japanese pharmaceutical firms. Strateg. Manage. J. 26, 121–140.

Polanyi, M., 1958. Personal Knowledge, Towards a Post Critical Epistemology. University of Chicago, Chicago, IL.

Powell, W., 1998. Learning from collaboration: knowledge and networks in the biotechnology and pharmaceutical industries. Calif. Manage. Rev. 40 (3), 228–240.

Ravichandran, T., Han, S., Mithas, S., 2017. Mitigating diminishing returns to R&D: the role of information technology in innovation. Inform. Syst. Res. 28 (4), 812–827.

Roach, M., Cohen, W.M., 2013. Lens of prism? Patent citations as a measure of knowledge flows from public research. Manage. Sci. 59 (2), 504–525.

Roe, David., 2018. Why Email Remains the Top Enterprise Collaboration Tool. Available at. https://www.cmswire.com/digital-workplace/why-email-remains-the-top-enterprise-collaboration-tool/.

Romer, P., 2008. Economic Growth. Library of Economics and Liberty. Available at. http://www.econlib.org/library/Enc/EconomicGrowth.html.

Rosenblat, T.S., Mobius, M.M., 2004. Getting closer or drifting apart? Q. J. Econ. 119 (3), 971–1009.

Schilling, M.A., Green, E., 2011. Recombinant search and the breakthrough idea generation: an analysis of high impact papers in the social sciences. Res. Policy 40, 1321–1331.

Singh, J., 2005. Collaborative networks as determinants of knowledge diffusion patterns. Manage. Sci. 51 (5), 756–770.

Singh, J., 2008. Distributed R&D, cross-regional knowledge integration, and quality of innovative output. Res. Policy 37, 77–96.

Spender, J.-C., 1996. Making knowledge the basis of a dynamic theory of the firm. Strateg. Manage. J. 17 (S2), 45–62.

Srikanth, K., Puranam, P., 2011. Integrating distributed work: comparing task design, communication, and tacit coordination mechanisms. Strateg. Manage. J. 32 (8), 849–875.

Szulanski, G., 1996. Exploring internal stickiness: impediments to the transfer of best practice within the firm. Strateg. Manage. J. 17 (S2), 27–43.

Teece, D., 1977. Technology transfer by multinational firms: the resource cost of transferring technological know-how. Econ. J. 87 (346), 242–261.

Thompson, P., Fox-Kean, M., 2005. Patent citations and the geography of knowledge spillovers: a reassessment. Am. Econ. Rev. 95 (1), 450–460.

Uzzi, B., Mukherjee, S., Stringer, M., Jones, B., 2013. Atypical combinations and scientific impact. Science 342 (6157), 468–472.

Van Alstyne, M., Brynjolfsson, E., 1996. Could the Internet Balkanize Science? Science 274 (5292), 1479–1480.

Van Alstyne, M., Brynjolfsson, E., 2005. Global village or cyber-balkans? Modeling and measuring the integration of electronic communities. Manage. Sci. 51 (6), 851–868.

Van Zeebroeck, N., van Pottelsberghe, B., Han, W., 2006. Issues in measuring the degree of technological specialisation with patent data. Scientometrics 66 (3), 481–492.

Walsh, J.P., Kucker, S., Maloney, N.G., Gabbay, S., 2000. Connecting minds: computer-mediated communication and scientific work. J. Assoc. Inf. Sci. Technol. 51 (14), 1295–1305.

Weitzman, M.L., 1998. Recombinant growth. Q. J. Econ. 113 (2), 331–360.

Wu, L., Hitt, L., Lu, B., 2016. Data Analytics Skills, Innovation, and Firm Productivity. The Wharton School Research Paper, pp. 86. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2744789.

Wu, L., Lou, B., Hitt, L., 2018. Data Analytics Supports Decentralized Innovation Communities. Working Paper. University of Pennsylvania.

Zhao, M., 2006. Conducting R&D in countries with weak intellectual property rights protection. Manage. Sci. 52 (8), 1185–1199.