Opportunity & Inclusive Growth Institute
FEDERAL RESERVE BANK *of* MINNEAPOLIS

The Paper Trail of Knowledge Spillovers: Evidence from Patent Interferences

Ina Ganguli
University of Massachusetts – Amherst

Jeffrey Lin
Federal Reserve Bank of Philadelphia

Nicholas Reynolds
Brown University

System Working Paper 18-06
February 2018

This paper was originally published as *Working Paper No. 17-44* by the Federal Reserve Bank of Philadelphia. This paper may be revised. The most current version is available at https://www.philadelphiafed.org/-/media/research-and-data/publications/working-papers/2017/wp17-44.pdf.

Opportunity and Inclusive Growth Institute
Federal Reserve Bank of Minneapolis • 90 Hennepin Avenue • Minneapolis, MN 55480-0291
https://www.minneapolisfed.org/institute

WORKING PAPERS

RESEARCH DEPARTMENT

WORKING PAPER NO. 17-44
THE PAPER TRAIL OF KNOWLEDGE SPILLOVERS: EVIDENCE FROM PATENT INTERFERENCES

Ina Ganguli
University of Massachusetts–Amherst

Jeffrey Lin
Research Department
Federal Reserve Bank of Philadelphia

Nicholas Reynolds
Brown University

December 2017

RESEARCH DEPARTMENT, FEDERAL RESERVE BANK OF PHILADELPHIA

Ten Independence Mall, Philadelphia, PA 19106-1574 • www.philadelphiafed.org/research-and-data/

The Paper Trail of Knowledge Spillovers: Evidence from Patent Interferences*

Ina Ganguli† Jeffrey Lin‡ Nicholas Reynolds§

December 2017

Abstract

We show evidence of localized knowledge spillovers using a new database of multiple invention from U.S. patent interferences terminated between 1998 and 2014. Patent interferences resulted when two or more independent parties simultaneously submitted identical claims of invention to the U.S. Patent Office. Following the idea that inventors of identical inventions share common knowledge inputs, interferences provide a new method for measuring spillovers of tacit knowledge compared with existing (and noisy) measures such as citation links. Using matched pairs of inventors to control for other factors contributing to the geography of invention and distance-based methods, we find that interfering inventor pairs are 1.4 to 4 times more likely to live in the same city or region. These results are not driven exclusively by observed social ties among interfering inventor pairs. Interfering inventors are also more geographically concentrated than inventors who cite the same prior patent. Our results emphasize geographic distance as a barrier to *tacit* knowledge flows.

Keywords: Localized knowledge spillovers, multiple invention, patents, interferences
JEL classification: O30, R12

*We thank Marcus Berliant, Jerry Carlino, Bob Hunt, and John Stevens, and conference and workshop participants at the North American meetings of the Urban Economics Association, the Federal Reserve System Committee on Regional Analysis, and the NBER Productivity Workshop for comments and suggestions, and Aaron Rosenbaum for excellent research assistance.

†University of Massachusetts Amherst, `iganguli@econs.umass.edu`.
‡Federal Reserve Bank of Philadelphia, `jeff.lin@phil.frb.org`.
§Brown University, `nicholas_reynolds@brown.edu`.

1 Introduction

Innovative activity happens in cities. Patenting, research and development laboratories, venture capital investments, new product introductions, and new activities are concentrated in large, densely populated places (Carlino and Kerr, 2015; Carlino et al., 2007 and 2017; Chatterji et al., 2014; Feldman and Audretsch, 1999; Lin, 2011). Yet these places also feature higher labor costs and prices for factors such as land (Rosenthal and Strange, 2004). Why then do firms and inventors choose to locate in dense, costly areas? Evidently, dense agglomerations offer some benefits for the creation or application of new knowledge.

One intriguing hypothesis is that inventors benefit from localized knowledge spillovers, or the external exchange of ideas among geographically proximate inventors and firms. This idea dates at least to Marshall (1920): "The mysteries of the trade become no mysteries; but are as it were in the air." Since Marshall, localized knowledge spillovers have featured in theories of economic growth and of cities. As Lucas (1988) noted, "what can people be paying Manhattan or downtown Chicago rents *for*, if not for being near other people?" But despite their theoretical and anecdotal significance, evidence of their existence remains mixed. There are at least two identification challenges. One, knowledge spillovers are hard to measure: "They leave no paper trail by which they may be measured or tracked" (Krugman, 1991). Two, alternative sources of agglomeration economies often yield similar predictions for measured productivity, wages, or other aggregates; i.e., different theories of the spatial concentration of firms and population are "observationally equivalent" (Duranton and Puga, 2004; Audretsch and Feldman, 2004).

We provide new evidence of localized knowledge spillovers that addresses these twin challenges. First, we construct a novel database of *patent interferences* terminated between 1998 and 2014—*simultaneous* instances of *identical* invention by two or more *independent* parties. Until 2013, the U.S. had a "first to invent" rule for assigning priority of invention, versus the "first to file" rule more common in the rest of the world and prevailing in the U.S. today. When the U.S. Patent and Trademark Office (USPTO) received applications with identical claims from multiple independent parties at roughly the same time, it was obliged to investigate the competing claims to determine which party was entitled to patent protection. This investigation, in the form of a patent interference proceeding, determined who had first conceived of the invention and reduced it to practice.

By recording instances of common invention, patent interferences create a unique record of common knowledge inputs. In the spirit of Weitzman (1998), the view that new ideas result from combinations of existing ideas suggests that interfering inventors share the same existing knowledge. For example, interfering inventors may have similar background in chemistry, or they may have similar information about market conditions. Thus, by examining the "paper trail" provided by patent interferences, we can identify common knowledge *inputs* in invention while circumventing the requirement of exactly measuring flows of knowledge across inventors. Moreover, common invention suggests a larger range of common knowledge inputs than those recorded in citations—especially shared tacit, or difficult to codify (and measure), knowledge inputs. For example, we discuss evidence from a 1981 interference case in which two research teams claimed a novel method of producing the hepatitis B antigen from yeast. Both teams had a range of common knowledge inputs, including an earlier advance awarded the 1976 Nobel Prize in medicine, previous attempts to produce the antigen using bacteria, and successes in producing other proteins using yeast. Interestingly, despite identical invention, the interfering teams only partially overlap in their citations of patent and non-patent literature. This pattern holds for our sample overall, suggesting that citations typically used to measure knowledge spillovers are likely to be noisy.

In our analysis, we show that interfering inventors tend to be geographically concentrated, consistent with the hypothesis that geographically proximate inventors are more likely to share common knowledge inputs. Other factors may also contribute to the geography of invention. In order to separately identify localized knowledge spillovers, Jaffe, Trajtenberg, and Henderson (1993) suggested comparing citation-linked inventor pairs with matched control pairs. We follow Jaffe et al. (1993) and construct matched control pairs to serve as a counterfactual. Control inventor pairs share a detailed technology classification and date of patent application, but compared with interfering inventor pairs, control inventors are "farther apart" in idea space. (This is a strong test since *none* of the agglomeration of control pairs is attributed to localized knowledge spillovers.) We find that interfering inventor pairs are 1.4 to 4.0 times more likely to live in the same city or region compared with control inventor pairs. In addition, we avoid scale and border problems by using distance-based tests of localization instead of administrative spatial units of arbitrary size (Duranton and Overman, 2005; Murata et al., 2014).

Our results do not appear to be driven exclusively by observed social ties. While interfering inventors are more likely to be connected by co-inventor ties, conditioned on these ties, interfering inventor pairs are still more localized compared with control pairs. Finally, we show that interfering inventors are more geographically concentrated compared with citation-linked inventors, pointing to the important role of geographic proximity in facililating *tacit* knowledge flows.

This paper makes several contributions. First, we contribute to the large literature on the sources of agglomeration economies (see surveys by Duranton and Puga, 2004; Rosenthal and Strange, 2004; and Combes and Gobillon, 2015). A key challenge to interpreting this evidence is that many other factors might also encourage inventors and firms to locate near one another. For example, firms might benefit from better matching with specialized workers, or they might better exploit opportunities for sharing other production inputs. Or skilled inventors may be attracted by superior amenities offered by big cities. Our evidence for knowledge spillovers, by focusing on the paper trail of inventions left by patents and interferences is more direct, and therefore avoids the identification problem of observational equivalence.

Second, we contribute to the literature that has provided evidence of localized knowledge spillovers using patent citations. Jaffe, Trajtenberg, and Henderson (1993) proposed a control-matching strategy to deal with the problem of inference under multiple sources of agglomeration economies. They argued that knowledge flows could be tracked via patent citations—when one inventor cited important prior art as a knowledge input. Then, they compared the geographic distance between a "citing" and "cited" inventor to the geographic distance between the cited inventor and a matched "control" inventor that was similar to the citing inventor in terms of technological classification and date of application, except for the citation link. Thus, the control patent-cited patent link represents the expected proximity of inventors working in the same time period and research field, but not conditioned on a "knowledge spillover" (i.e., a citation). If the inventors of the citation-linked patent pair are observed in closer proximity than this benchmark, then this is strong evidence that a localized knowledge spillover has occurred, since we have now accounted for the underlying geographic distribution of research activity and hence why inventors might be located together. Jaffe et al. find evidence consistent with localized knowledge spillovers. However, subsequent work, including Thompson (2006) and Thompson and Fox-Kean (2005), found that these results were fragile—choosing to exclude patent examiner-added citations, for example, or selecting control patents in a different way, eliminated the Jaffe et al. result. Recently, Murata et al. (2014) proposed a distance-based version of the Jaffe et al. test based on methods developed by Duranton and Overman (2005), and they found evidence in support of localized patent citations.

In contrast, our evidence from interferences does not rely on the "noisy" signal of knowledge spillovers from citations. Jaffe, Trajtenberg, and Fogarty's (2000) survey of patenting inventors reported that "one-half of all citations do not correspond to any spillover," and Jaffe, Trajtenberg and Henderson (1993) acknowledged that "an enormous number of spillovers [occur] with no citations." (Inventors strategically not citing known prior art contributes to this problem.) Interferences sidestep these issues by avoiding the use of citations to measure knowledge spillovers. We find that interfering inventors do share significantly more backward citations compared with control inventor pairs. Yet we also document that interfering inventors only partially overlap in their citations, suggesting that citations are indeed a noisy measure of knowledge spillovers. Moreover, interferences may better capture spillovers of *tacit* knowledge, i.e., forms of knowledge that are difficult or impossible to codify into a citation. We find that interfering inventors are more geographically concentrated compared with citation-linked inventors, consistent with the Jaffe, Trajtenberg, and Henderson (1993) conjecture that citations are a *lower bound* on the strength of localized knowledge spillovers.

Finally, we contribute to the literature on multiple invention and discovery in science. Historians of science have noted the frequency of simultaneous, independent discovery—e.g., the independent invention of calculus by Newton and Leibniz, or the independent formulation of the theory of natural selection by Darwin and Wallace. Robert Merton (1973) called these instances "multiples." The commonplace nature of multiples has led some to speculate that they must be "in the air, products of the intellectual climate of a specific time and place" (Gladwell, 2008). Compared with earlier efforts to measure multiples (Ogburn and Thomas, 1922, Bikard, 2012), our database is both large and relies on the real-time declaration of multiples by patent examiners instead of *ex post* measurement by researchers. We also extend this literature to the use of multiples for the identification of localized knowledge spillovers.

2 Patent interferences

2.1 Background

Patent interferences were a unique feature of U.S. patent law. Through March 16, 2013, the U.S. had a "first to invent" rule for assigning priority of invention, versus a "first to file" rule more common in the rest of the world and that prevails in the U.S. today. When the USPTO received patent applications from multiple, independent parties with one or more identical claims at roughly the same time, it was obliged to investigate the competing claims to determine which party was entitled to patent protection. This investigation, in the form of a patent interference proceeding, determined who had first conceived of the invention and reduced it to practice. Typically, parties submitted dated laboratory notebooks, testimony by associates, and media reports as evidence of first invention. In this section, we summarize several key institutional features of patent interferences.[1]

An interference was declared by a patent examiner, during their routine search for prior art, when (i) at least two simultaneous U.S. patent applications or (ii) one U.S. patent application and a recently-issued patent contained identical claims of invention. The claim(s) of invention must have satisfied standard patentability rules—i.e., the claims must have been in an patent-eligible class, useful, novel, and non-obvious. In addition, the USPTO required that a timing rule be satisfied in order to avoid interferences resulting from the disclosure of patent applications themselves (i.e., publicized patent applications leading to copycat inventions). Thus, in the case of two or more interfering applications, the dates of application must have been no more than 3 months apart. In the case of an interfering issued patent and pending application, (a) the application's date must have been more than one year before the patent's grant date and (b) the application's date must have been no less than 3 months after the patent's application date.[2]

Upon declaring an interference, the examiner defined "counts" corresponding to the interference. Each count was characterized by a distinct invention at issue; each application might claim several distinct inventions so an interference might involve multiple counts. The case was then sent for hearing before a rotating three-judge panel from the Board of Interferences. Inventors were assigned a "benefit date"—typically, the date of application, either at the USPTO or a foreign patent office. The inventor with the earlier benefit date was referred to as the "senior party". The burden of proof—i.e., demonstration of an earlier conception and reduction to practice—was on the "junior party". Interfering inventors typically submitted lab notebooks, eyewitness testimony, and other forms of independent corroboration to prove they were the first to invent.

Interference cases were terminated by the judges' decision on priority or for some other reason. Table 1 summarizes common interference dispositions. A decision on priority meant a judgment that one party had first conceived of the invention and reduced it to practice.

[1] More details about the patent interference proceedings can be found in Calvert (1980), Calvert and Sofocleous (1982), Cohen and Ishii (2006), de Simone, Gambrell and Gareau (1963), and Kingston (2001). Lin (2014) reviews this literature and provides summary statistics for the patent interferences used in this study.

[2] Specifically, interfering claims among pending applications must be made within 1 year of each other (35 U.S.C. 135.b.2). In cases where an application's claims interfere with an already-issued patent, the claims must be made no later than 1 year prior to the patent's issue date (35 U.S.C. 135.b.1), and typically no later than 3 months after the patent's original application date (37 C.F.R. 1.608).

Disposition	Description or example	Patent(s) to
Priority	Y conceived and reduced to practice first	Y
Settlement	X concedes; settlement terms confidential	Y
Abandonment	X concedes; no settlement	Y
Concealment	X kept invention secret, concealed best mode, or delayed filing	Y
Derivation	X stole invention from Y; i.e., prior conception by Y and communication of conception to X	Y
No interference in fact	Claims are actually distinct	X and Y
Common ownership	X and Y work for same conglomerate	X or Y
Unpatentable	Claims are not patentable (e.g., obvious)	No one

Table 1: Common case dispositions for interference between inventors X and Y

Parties could also settle at any stage. Normally, details of these agreements were kept secret. One party might also concede the case, without a settlement taking place—for example, if they realized their case was weak.

Other potential termination types are useful for distinguishing instances when inventors shared knowledge inputs versus other reasons. For example, according to interference rules, an interfering inventor would lose priority if the inventor had not immediately filed for a patent application following conception and reduction to practice.[3] In particular, the disclosure of the timeline of invention was the primary purpose of the interference proceeding. The alternative to admitting an intentional delay in applying for a patent would be to concede a later date for reduction to practice, weakening the case for priority.[4]

Sometimes, the Board of Interferences might decide that the examiner that declared the interference was mistaken, and that there was no interference in fact. This reported result is important because it allows us to isolate true multiples, and not just near misses.

[3]Cohen and Ishii (2006) argue that interferences correspond to an incumbent-entrant game where incumbents decide to keep inventions secret for some period of time before filing a patent application. The requirement to promptly disclose an invention may have reduced the relevance of this margin as decisions went against interfering inventors who chose to keep their inventions secret for some time.

[4]Interference cases appear to vary in terms of whether inventors are aware of each other's efforts. In Lutzker v. Plet (1988), the United States Court of Appeals, Federal Circuit, affirmed that Lutzker was not entitled to a patent for a canape maker, despite having established conception and reduction to practice in early 1976, since he had delayed disclosure and filing for a patent until late 1980. In contrast, Plet received priority by demonstrating her conception and reduction to practice by early 1980, with a filing date of March 3, 1980. The original decision by the Board of Interferences cited the failure of Lutzker to show renewed activity towards disclosure "until after Plet entered the field" as an important factor in the judgment against him.

The USPTO also made sure that interfering inventors did not share other factors. Claims from parties working for the same firm (e.g., different branches of a large corporation) were dismissed. If one party's application was derived—i.e., the invention was disclosed and communicated to them—then that was grounds for an adverse judgment. For example, an invention based on a stolen idea was considered derived. Finally, in some less common outcomes, the board ruled that the interference count was anticipated or otherwise unpatentable. As with no-interference-in-fact, these judgments could be interpreted as mistakes by the original patent examiner.

Several features of interference practice allow us to rule out alternative explanations—outside of shared knowledge inputs—for multiple invention. First, interferences were declared by a patent examiner who specialized in a particular technological area. Thus, interfering claims were likely to be detected. In some cases, the examiner was alerted to a possible interference by one of the applicants, but an interference is distinct from patent infringement, in which the holder of an existing patent sues an infringing party. In contrast to infringements, private parties could not sue for an interference.

Second, interferences must involve parties with simultaneous pending applications for patents. This is important for several reasons. One, this feature makes interferences distinct from patent infringements, which typically involve leaders and followers. Two, the simultaneity requirement reduces the likelihood of copying or stealing—that is, that one inventor's claims are directly sourced from the competing party. Instead, it is more likely that both parties are drawing knowledge inputs from shared information. In addition, as mentioned earlier, evidence of stealing or espionage is grounds for adverse judgment in the interference decision.

Third, to the extent that interference cases are costly to prosecute, interferences are likely to involve valuable patents, and thus actual inventions.[5] Fourth, interferences between parties with common ownership interests were not allowed. Thus, they seem unlikely to result from other shared factors in common or from within-firm spillovers. Fifth, as mentioned earlier, no-interferences-in-fact help us to distinguish between identical inventions and near misses. Sixth, some interferences represent cases where patent applications completely overlap. Thus, *contra* Schmookler (1966), interferences can identify identical inventions, versus

[5]A recent interference case decided in February 2017 involved patent rights to Crispr, a powerful gene-editing technique. (The decision date puts the case outside our sample.) In that decision, the Board of Interferences found no interference in fact—the inventions claimed by the competing inventors, assigned to the Broad Institute and the University of California, were separate and did not overlap. The validity of the Broad patents was a "surprise" to researchers in the field; as a sign of the value of the invention, the stock of the licensee to the Broad patents went up sharply following the decision (Pollack, 2017).

near misses. The USPTO tracks the number of application claims corresponding to each interference count. Therefore, it is possible to separate interferences where all application claims are in interference from other cases where only some application claims are in interference.

2.2 Case study

One interference (number 102,416) involved competing claims for the method of producing the hepatitis B vaccine from yeast. This case is somewhat unusual in that the patent was especially valuable and appeals dragged through the court system for many years. However, several features of the case highlight important general features of interferences as indicators for shared knowledge inputs.

The case involved two competing teams. Party 1 included William Rutter, Pablo Valenzuela, Benjamin Hall, and Gustav Ammerer, who were scientists at the University of California and University of Washington and whose work was in part funded by Merck. Party 2 included Ronald Hitzeman, Arthur Levinson, and Daniel Yansura, who were scientists at Genentech, based in South San Francisco. Rutter et al. filed their application on August 4, 1981, claiming simultaneous conception and reduction to practice on June 30, 1981. Hitzeman et al. filed their application on August 31, 1981, claiming conception on February 3, 1981 (five months before Rutter et al.) and reduction to practice on July 20 (three weeks after Rutter et al.). After a lengthy set of motions and appeals, the case was finally decided by the U.S. Court of Appeals for the Federal Circuit (2001) in favor of Rutter et al., on the basis that Hitzeman et al.'s earlier claimed conception date was invalid. The court ruled that Hitzeman et al.'s actual conception date was July 20, 1981. Since Rutter's team now had an earlier conception and reduction to practice, they were awarded the patent.

Both teams had knowledge inputs in common, according to Kleid's (2002) oral history of the case and the decision by the appeals court (2001). It was already known that the hepatitis B antigen could be purified from the blood of certain infected humans. This so-called "Australia antigen" could then be used as a vaccine; this advance was awarded the 1976 Nobel Prize in Medicine, but the vaccine produced in this way remained costly to manufacture. Many teams, including the two in interference, speculated that a hepatitis B antigen might be produced using bacteria or yeast. However, previous attempts by other scientists to use E. coli to produce the antigen had failed.

Thus, members of both teams had similar knowledge inputs. Among these common knowledge inputs were the market need for a low-cost hepatitis B vaccine; the costly "Australia antigen" method; the failed attempts using E. coli; and the successful use of yeast to produce other proteins. Interestingly, the published patents indicate that the interfering teams only partially overlap in their citations of patent and non-patent literature. The patents of the Hitzeman and Rutter teams cite 6 and 9 prior patents, respectively, but only 4 of these citations appear on the the patents of both teams. Of the 19 and 34 prior scientific articles cited by the patents of the respective teams, only 6 are shared.[6] These patterns suggest that citations are noisy measures of shared knowledge.

Notably, the interfering teams were geographically co-located and linked by previous co-inventor ties. Rutter and Valenzuela both lived in San Francisco, while Hitzeman, Levinson, and Yansura lived nearby, in Pacifica, Burlingame, and San Francisco, respectively. In the late 1970s and in 1980, Hitzeman had collaborated with scientist John Carbon to successfully produce the "interferon" protein in yeast, providing a model for producing the hepatitis B antigen in yeast. After this work, he joined Genentech and collaborated with Hall and Ammerer on other projects involving interferon. Shortly thereafter, both teams decided to try using yeast to manufacture the antigen. We will return to the role of co-inventor ties in interfering parties below.

3 Data and methodology

3.1 Interferences, patents, and applications

We constructed a database of patent interference cases for our analysis. Our database starts with information from the 1,329 interference *decisions* issued by the U.S. Board of Interferences between 1998 and 2014.[7] These decisions were downloaded from the Board's "e-FOIA Reading Room." From each decision, we record information about the *case*, the *parties*, the *application(s)* and/or *patent(s)*, the *claims*, and the *inventors*. We note the following: (1) each case is argued between two or more parties; (2) each party may have one or more (co-)inventors; (3) each party may also have one or more applications and/or patents in interference; (4) each application or patent makes one or more claims; (5) one or

[6] The above comparison excludes patents and articles dating from after the August 1981 application dates of the Hitzeman and Rutter patents. The Hitzeman patent was issued in 1989 and eventually invalidated, while the Rutter patent was not issued until 2002. The sole examiner-added citation in the Rutter patent was to the Hitzeman 1989 patent.

[7] There are a few decisions related to interferences declared well before 1998, as far back as the early 1980s, including the hepatitis B case. On average, however, the lag between interference declaration and decision dates is a few years. See Calvert and Sofocleous, 1982, 1986, 1989, 1992, and 1995.)

more of these claims are declared by the examiner to be in interference.

The decisions typically report the: (i) names of the interfering inventors; (ii) seniority status of each party; (iii) associated patent and application numbers; (iv) assignees; (v) judges' names; (vi) application claims in interference; (vii) decision on priority at the claim level (if there was one) or other disposition of the case; (viii) legal counsel; and (ix) hearing and decision dates. Sometimes, terse decisions omit some of these details. When available, we collect these details using additional documents found on the USPTO's "eFile" site or the Patent Application Information Retrieval (PAIR) service. The eFile site sometimes lists the notice declaring the interference, from which we can observe (x) inventors' location of residence. For cases with documents available on the eFile site, we also record notices of settlement agreements. These notices acknowledge the existence of a settlement agreement, as opposed to a decision on priority or some other outcome.[8] The PAIR service provides an alternate source of information on assignees, case disposition, the decision date, and inventors' locations. Note that for inventors never (eventually) issued a patent, information on inventor location is *only* available on the notice of interference on the eFile site or the PAIR service.

Table 2 summarizes case dispositions for our sample. The first two columns display frequencies and the share of cases by disposition for our full sample. Nearly 20 percent of cases resulted in a judgment on priority, while nearly 60 percent of cases were conceded. Concessions occur when one party files a request for adverse judgment. An abandonment occurs when one party fails to file at some stage of the case. We code these outcomes as they are noted in the decisions. Absent detail in the decision, it is difficult to ascertain the motivations for concessions (and we cannot rule out abandonments if failure to file is not mentioned.) However, for the sub-sample of 977 cases that we are able to match to documents on the eFile site, we code cases including an acknowledgement of settlement. In this sub-sample, settlements constitute the majority of concessions and nearly one-third of all cases. We are unable to characterize the remaining conceded cases that have no acknowledgement of settlement or text in the decisions referring to a failure to file.

Some of our analysis focuses on interference cases where the board's decisions report a settlement or judgment on priority. Shared knowledge inputs are more likely in these cases compared with other case dispositions. For example, cases dismissed for "no interference in fact" seem less likely to involve exactly common knowledge inputs. The frequencies of

[8]Unfortunately, settlement agreements are sealed. Thus, we can note their existence, but we cannot analyze their contents.

other dispositions are listed in the bottom half of Table 2. About nine percent of cases were dismissed because the claims were deemed unpatentable. Five percent were dismissed because the interfering parties were discovered to have assigned rights to a common owner, e.g., a multinational firm.[9] Three percent of cases were dismissed after a finding of no interference in fact.

Disposition	*Full sample*		*eFile*
Number of cases	1,329		977
Decision on priority	260	19.6%	19.7%
Conceded, total	781	58.8	58.1
. . . settled	.	.	32.8
. . . abandoned	92	6.9	5.5
. . . all other reasons	.	.	19.8
No interference in fact	46	3.5	3.4
Common ownership	64	4.8	4.7
Unpatentable	122	9.2	9.6
Other	56	4.2	4.5

Table 2: Distribution of interference case dispositions

Table 3 displays statistics on claims of invention by parties according to their seniority status. (Recall that seniority status is typically determined by first-to-file and that the burden of proof in showing earlier conception and reduction to practice is on the junior party.) Panel A shows statistics for interference cases where we are able to observe both the claims made on all involved applications and the claims in interference. Panel B shows statistics for interference cases where we are able to observe only the claims in interference.

Interference counts correspond to most of application claims. The first row of Panel A shows that junior parties tend to make about 3 more claims on average compared with senior parties. (This difference in means is significant at the 5 percent level.) Of the 26 and 23 claims, respectively, made by junior and senior parties on average, 20 and 19 claims are declared in interference. (For the sample for which we only observe interfering claims, slightly fewer claims, 18 and 17, are declared in interference.)

Senior parties tend to win two-thirds of the time. For the sample described in Panel A, junior parties lose 13 out of the 20 claims in interference, or 65 percent. Partial decisions

[9] In at least one case, a merger appeared to have been *caused* by the pending interference.

are rare. Out of the 688 junior parties, 465 (68 percent) lost *all* of their claims declared in interference. Out of the 687 senior parties, 221 (32 percent) lost all of their claims declared in interference. (The difference in these rates is significant at the 1 percent level.) These rates are similar to the larger sample described in Panel B (for which we observe only claims in interference). Overall, the patterns are also similar to those seen in the subsamples of cases decided on priority or cases conceded.

Next, we match these data on interference cases to patent and application data in order to identify the location for inventors and to identify co-located inventors. If an interference is decided against a party without an issued patent, then the losing party's patent application is never passed for issue. Since most *patent* databases only include data for issued patents, we must obtain information about denied applications from the PAIR service, which contains application-level data for applications filed since 2001. We extract the name and location of the first-named inventor, the identities of the patent examiner and attorney, and the filing date for each application. We also take advantage of the associated transaction history file to record all interference cases associated with the patent application, decision dates, and whether the ruling was favorable or unfavorable.[10]

For issued patents, we also match our interference case database to the inventor disambiguation dataset of Lai et. al (2013). For all patents issued by the USPTO between 1975 and 2010, this includes the names and locations of all inventors, the patent application date, and assignee. This dataset also includes a unique inventor identifier that is consistent across all patents, as a result of a name disambiguation algorithm. Importantly, this allows for the examination of the entire inventing careers of particular inventors and for the construction of a network of co-inventor ties that allows for the measurement of network distance.

The importance of supplementing the decisions database with the eFile and PAIR files is illustrated by the frequency of inventor locations by data source reported in Table 4. While 75 percent of inventors are located in the Lai et al. (2013) database, the remaining 25 percent of inventors' locations are recorded *only* in the decisions, eFile, or PAIR. Across cases, we are able to record inventor locations for nearly 88 percent of interference cases (Table 5).

Finally, for matching using technology classification information, we obtain technology classification information directly from the USPTO's Master Classification File. This information is available for all issued patents, but only for patent applications filed in 2001 and later. For earlier patent applications, we instead obtain classification information from the

[10]In nearly every case where they overlap, information available on the PAIR record confirms information recorded from the decision.

	All cases		*Priority decisions*		*Conceded*	
	Jr.	Sr.	Jr.	Sr.	Jr.	Sr.
A. Application claims observed						
Claims in application(s)	26.0	22.8^b	26.7	24.1	24.7	21.1^b
and/or patent(s)	(26.9)	(26.6)	(35.9)	(31.4)	(22.0)	(24.1)
Claims in	20.0	18.8	21.0	19.3	18.1	17.6
interference	(21.5)	(22.5)	(21.5)	(24.4)	(16.8)	(20.8)
Claims lost	13.3	6.6^c	17.9	3.9^c	12.8	5.0^c
in decision	(19.4)	(16.5)	(20.6)	(12.5)	(17.4)	(11.1)
Number of parties	688	687	121	116	358	360
Lost all appl./pat. claims	276	173	73	14	131	89
Lost all interfering claims	465	221	101	20	251	114
B. Only interfering claims observed						
Claims in	17.6	17.2	18.8	17.7	16.4	16.5
interference	(18.9)	(20.0)	(19.8)	(20.5)	(15.3)	(18.8)
Claims lost	12.7	6.5^c	15.9	4.1^c	12.0	6.0^c
in decision	(17.3)	(15.2)	(19.5)	(11.7)	(15.3)	(18.8)
Number of parties	1,236	1,102	257	214	627	563
Lost all interfering claims	910	405	212	47	469	221

Table 3: Claims for Senior and Junior Parties.

This table reports means and standard deviations (in parentheses) for senior and junior parties. Seniority is determined before an interference proceeding begins, according to the earliest benefit date on file. (Typically, the benefit date is the earliest date of application to the USPTO or a foreign patent authority.) Number of claims is the sum of claims across all applications filed by each independent interfering party. H_0: Difference in means by seniority is zero. ^{a}—$p < 0.10$; ^{b}—$p < 0.05$; ^{c}—$p < 0.01$.

Lai et al. (2013)	5,256	75.3%
PAIR	664	9.5 %
Decisions or eFile	1,056	15.1%
Total	6,976	100%

Table 4: Frequency of inventors by source of location information

All inventor locations recorded	1,126	84.8%
At least 1 inventor location for each party	36	2.7
Missing inventor location for at least 1 party	167	12.6
Total	1,329	100%

Table 5: Frequency of interference cases by availability of inventor location

PAIR service, which only records the primary classification.

For inventors with issued patents, we obtain the latitude and longitude of each inventor from the database of Lai et al. (2013). For inventors with only patent applications, we match the place and state of the inventor from either the PAIR service or the eFile notice of interference to the Census Gazetteer file, which maps place names to latitude and longitude. We then compute the geodesic distance for each possible pairing of inventors within an interference case. For an interference between a party with m co-inventors and another party with n co-inventors, there are mn pairwise combinations of inventors. We record the median, mean, and minimum distance for each interfering pair of parties. (Also, following the convention of earlier work, we record the distance between the first-named inventors of each patent.) We report results using minimum pairwise distances, but results using alternative measures are similar.

3.2 Control patents and simulation of counterfactual distances

Can localized knowledge spillovers be identified by the geographic concentration of inventive activity? Inventors might choose locations with other important factors (e.g., capital, skilled workers) or greater demand. Thus, it is unclear whether the co-location of inventors indicates that knowledge flows are empirically relevant. Jaffe et al. (1993) suggested that by comparing the geographic distribution of citation-linked inventors to appropriately matched control patents, they could separately identify these spillovers.

We test for tacit knowledge flows via geographic and social proximity of interfering inventors, following the control-matching strategy of Jaffe et al. (1993). We compare interfering inventor pairs to *control* patent pairs—where a control pair includes one interfering application and one issued control patent matched on technology class and application date. The idea is that the spatial distribution of control patent pairs is an appropriate counterfactual to the observed spatial distribution of interfering inventors. That is, save for the fact that control inventor pairs did not invent identically, they face a similar location choice problem compared with interfering inventor pairs. In constructing this counterfactual, we hope to control for all factors except common knowledge inputs. As Jaffe et al. (1993) note, this is a conservative estimate in the sense that only the concentration *in excess* of the control pairs is attributed to localized knowledge spillovers.

Following Thompson and Fox-Kean (2005), we use both 3- and 6-digit technology classifications to select control patents. We also use distance-based methods following Duranton and Overman (2005) and Murata et al. (2014) to overcome scale and border problems in using administrative spatial units (e.g., municipalities) of arbitrary size.

Since some interference cases involve more than two parties, our database of 1,329 interference cases involves 1,401 interfering pairs of inventors.[11] We select a set of control patents that are similar to the invention described by each party's patent(s) or application(s) declared in interference. First, we require a control patent to share a 3- or 6-digit technology classification with a party's patent(s) or application(s). Second, control patents must have an application date within 180 days of of the application date of an interfering application. We are able to obtain a set of suitable control patents for nearly every interfering inventor pair—only 24 pairs lack suitable 3-digit controls and 32 pairs lack suitable 6-digit controls (Table 6).

Next, each of these control patents is now eligible to form a *control pair*. A control pair matches a control patent to an interfering patent or application. Control patents are matched to the opposing party's interfering application(s). This matching structure is identical to Jaffe et al. (1993). In their application, *cited* patents are used to identify control patents. Control patents are then matched to *citing* patents.

In our sample, the interfering application(s) of the first party are used to identify control patents. Control patents are then matched to the interfering application(s) of the second party. Table 6 shows that the pool of control patents is large. Conditioned on finding a control, the average interfering inventor pair is associated with 706 control patents matched on the 6-digit technology class and 6,457 controls matched on the 3-digit technology class.

We estimate sample statistics using our database of interfering pairs. To infer the presence of localized knowledge spillovers, we compare these to simulated distributions of control pairs, following Duranton and Overman (2005) and Murata et al. (2014). First, we randomly select a control pair from the set of permissible control pairs. With this random draw of control pairs, we can estimate kernel density functions for the distribution of pairwise geographic

[11]Twenty-four cases involve three parties, three cases involve four parties, and one case involves five independent parties.

Number of interfering pairs*			1,401
A. 3-digit controls†			
Interfering pairs w/o eligible 3-digit controls			24
	Control patents	*Control pairs*	
Minimum	1	1	
Median	5,011	5,516	
Mean	6,457	7,675	
Maximum	38,013	70,992	
B. 6-digit controls†			
Interfering pairs w/o eligible 6-digit controls			32
	Control patents	*Control pairs*	
Minimum	1	1	
Median	180	203	
Mean	706	820	
Maximum	8,188	20,435	

Table 6: Control patents and pairs

*—Interfering pairs are two independent parties, each with one or more applications or patents, in a declared interference.
†—*Control patents* match the technology class and application date (within 180 days) of an application or patent by an interfering party. They are then paired with an application or patent by the opposing interfering party to form a *control pair*.

distance. We also estimate a number of other simulated statistics.

We evaluate significant departures from our counterfactuals by repeating this exercise and drawing a new set of control pairs 1,000 times. Then, we rank the simulated kernel density estimates at 100 evenly spaced pairwise distances, and select the 50th ranked simulated kernel density at each distance to construct the lower 5 percent confidence band and the 950th ranked to construct the upper 95 percent confidence band. These *local* confidence bands only allow us to make local statements about the relationship between the interfering and counterfactual distributions (i.e., if the density of pairwise distances for interfering pairs is above the upper counterfactual band at 100 km we could say that interference pairs are localized at 100 km). By construction there is a 5 percent probability for each particular distance that a random draw of counterfactual pairs shows localization and thus across all distances a much higher likelihood of displaying localization at some distance. This effect is attenuated slightly by autocorrelation between the densities at different distances. However, to make global statements about localization, we construct global confidence bands. First, we define d_a as the median distance of all counterfactual pairs. Then we return to our simulated draws and search for an upper and a lower local confidence band such that, when we consider them across all distances between 0 and d_a, only 5 percent of our simulated K-densities hit them. In general, we end up selecting approximately the 99 percent local confidence band as the 95 percent upper global confidence band and the 1 percent local confidence band as the 5 percent lower global confidence band.

4 Results

4.1 Shared codified knowledge inputs

Table 7 shows summary statistics of shared codified knowledge inputs (citations and technology classes) for interfering applications/patents and control pairs. We present separate simulated statistics for 3- and 6-digit controls. First, interfering patents and applications tend to have a similar number of backwards citations compared with 3- and 6-digit control patents. Second, interfering patents and applications tend to have fewer U.S. patent classifications and subclassifications compared with 3-digit and 6-digit control patents. These differences are consistent with the rate of interfering applications—about a quarter—that are never passed for issue. For these applications, we only observe one classification and subclassification.

Interfering pairs tend to share more codified knowledge inputs compared with control

		3-digit controls		6-digit controls	
	Interfering	μ	90% C.I.	μ	90% C.I.
A. Means for applications/patents					
Backwards citations	11.1	11.8	(11.1, 12.7)	11.8	(11.0, 12.6)
USPC classes	1.93	2.28	(2.23, 2.34)	2.24	(2.18, 2.29)
USPC subclasses	4.93	5.95	(5.71, 6.19)	7.06	(6.75, 7.41)
B. Means for control pairs					
Backwards citations shared	3.53	0.03	(0.01, 0.06)	0.33	(0.20, 0.62)
USPC classes shared	1.25	0.85	(0.82, 0.87)	1.02	(0.99, 1.04)
USPC subclasses shared	1.50	0.10	(0.08, 0.13)	0.68	(0.64, 0.71)

Table 7: Interfering inventors share codified knowledge inputs

This table compares means of interfering patent pairs to simulated means, 5th- and 95th-percentile estimates for control pairs. A control pair includes 1 interfering application and 1 issued control patent that share a technology class and application date. Simulated CIs based on 1,000 random draws from eligible control pairs. The sample is cases with decisions on priority and concessions.

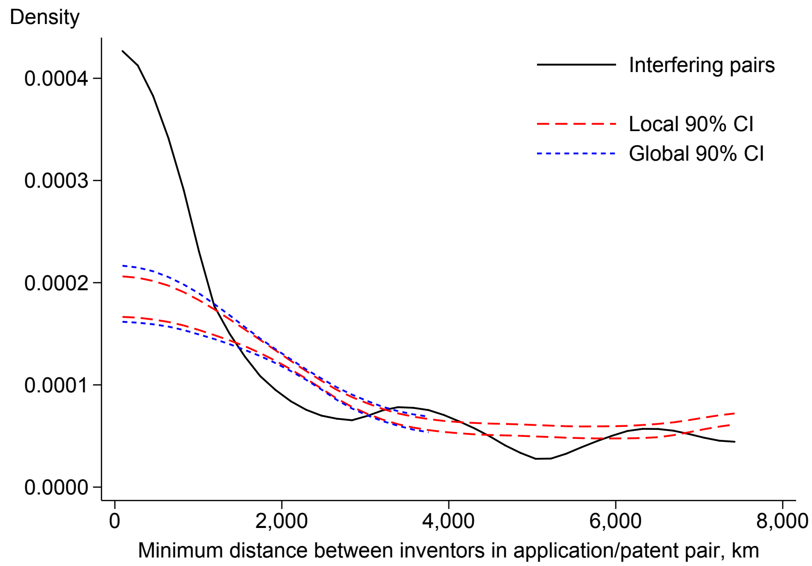
pairs. Panel B shows that interfering pairs share 3.5 backwards citations, compared with 3-digit control pairs that share 0.03 backwards citations and 6-digit control pairs that share 0.33 backwards citations. This is one indicator that interfering inventors are closely related in idea space. Interfering pairs also share more U.S. patent classifications. Interfering pairs tend to share 1.25 classes and 1.50 subclasses. This is more than the number of classes and subclasses shared by both 3- and 6-digit control pairs.[12]

4.2 Geographic proximity

We next show that interfering pairs of inventors are more localized compared with control inventor pairs not linked by identical invention. Figure 1 compares the density of geographic distance between pairs of interfering inventors with that of similar, non-interfering control pairs. The black line shows the estimated kernel density function for all interfering pairs for which we were able to find suitable controls. The dotted red lines show the local confidence bands for control pairs. The dotted blue lines show the global confidence bands for control pairs. For each interfering and control pair we used the minimum pairwise distance among co-inventors. In Panel A, we show results using 3-digit control pairs; in Panel B, we show results using 6-digit control pairs. Thus, while the observed distribution of pairwise distances

[12]Recall that a patent can have more than one classification, and that control patents are matched to the opposing party in interference. Hence, control pairs need not share a patent class.

Panel A. 3-digit controls



Panel B. 6-digit controls

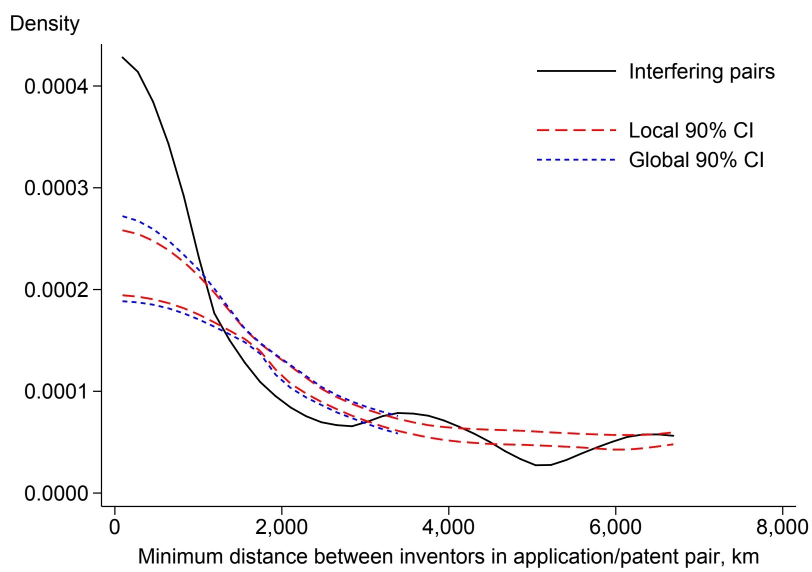


Figure 1: Density of geographic distance between inventor pairs

These graphs compare the density of geographic distances between pairs of interfering inventors to the distribution for similar non-interfering control pairs. The black line shows the estimated kernel density function for all interfering pairs for which we were able to find suitable controls. The dotted red lines show the local confidence bands of the kernel density for non-interfering control pairs. The dotted blue lines show the global confidence bands of the kernel density for non-interfering control pairs. For each interfering and control pair we use the minimum pairwise distance among co-inventors.

for interfering pairs is the same in both panels, the counterfactual distribution is more geographically concentrated for the 6-digit control pairs shown in Panel B. However, for either counterfactual, we find that interferences are indeed significantly geographically localized. This evidence is consistent with geographic proximity facilitating the sharing of common knowledge inputs.

The result that interfering inventors pairs are more localized is robust to conditioning on decision type. Table 8 shows average pairwise distances for interfering pairs and control pairs with simulated 90 percent confidence intervals. On average, interfering pairs in priority decisions and concessions are 3,500 km apart, compared with 4,800 km separating 3-digit control pairs of inventors and 4,400 km separating 6-digit control pairs of inventors.

		3-digit control		6-digit control	
	Interfering	μ	Sim. 90% C.I.	μ	Sim. 90% C.I.
Priority decisions and concessions	3,451	4,777	(4,563, 5,985)	4,425	(4,219, 4,623)
Priority decisions only	3,603	4,714	(4,321, 5,121)	4,281	(3,889, 4,688)

Table 8: Interfering inventor pairs are closer together (km)

	Same city			Within 161km (100mi)		
	Int.	3-dig.	6-dig.	Int.	3-dig.	6-dig.
Priority decisions and concessions	2.7%	0.8%*	2.0%	13.8%	5.2%*	8.2%*
Priority decisions only	2.8%	0.7%*	2.0%	11.7%	5.2%*	8.2%*

Table 9: Interfering inventor pairs are co-located

*—Difference compared with interfering pairs is statistically significant ($p < 0.10$).

Though interfering inventors are closer together on average, the average pairwise distances may obscure the relevant range of geographic proximity for localized knowledge spillovers. Table 9 presents results using geographic matching rates, as in Jaffe et al. (1993). Nearly 3 percent of interfering inventors are located in the same city, compared with 1 and 2 percent of 3- and 6-digit control pairs, respectively. (The difference compared with 3-digit control pairs is significant at the 90 percent level.) We also compare geographic matching rates for regions. This test is similar to the original matching-rate tests at the metropolitan area level reported by Jaffe et al. (1993), except that we leverage the micro-geography of inventor location compared with co-locating within the same county or set of counties. Between 12 and 14 percent of interfering inventor pairs are less than 161 km (100 miles) apart, compared with 5 to 8 percent of control inventor pairs. These differences are significant at the 90 percent level. In sum, interfering inventor pairs are 1.4 to 4.0 times more likely to locate in the same city or region compared with control inventor pairs.

4.3 Comparison to citation-linked controls

Are spillovers of *tacit* knowledge more localized than other forms of knowledge that are more easily codified? Arzhagi and Henderson's (2008) results suggest that the external benefits to advertising agencies in Manhattan attenuate quickly over space—they dissipate in as little as 750 meters. To the extent that interferences can capture spillovers of both tacit and codified knowledge, their localization could provide evidence that tacit knowledge spillovers require even closer proximity. To test this hypothesis, we compare the observed distribution of geographic proximity between interfering inventors to a particularly strong counterfactual: pairs of control patents and interfering patents linked by citation. Thus, our "control pairs" in this exercise are the "treated" pairs in Jaffe et al. (1993). For each interfering pair, we identify potential controls as patents *cited by* one of the interfering parties. Then, we form cited-citing control pairs by matching an interfering application to one of the these cited control patents.

Figure 2 shows this comparison between interfering inventor pairs and cited-citing control pairs. The black line again shows the observed density of pairwise distances between interfering inventors. (This is the same density reported in Figure 1.) The red and blue lines show simulated local and global 90 percent confidence intervals, respectively, for the density of pairwise distances between cited-citing control pairs. Even compared against the geography of citation-linked patents, interfering patents are significantly more localized. The result that interfering inventors are more geographically concentrated compared with citation-linked inventors is consistent with the Jaffe, Trajtenberg, and Henderson (1993) conjecture that citations are a *lower bound* on the strength of localized knowledge spillovers.

4.4 Matching-rate regressions

What information about inventors' knowledge is contained in interferences, beyond commonly used textual measures such as technology classification or citations? Is the geographic localization of interfering inventors driven by observed social network ties? In this section we

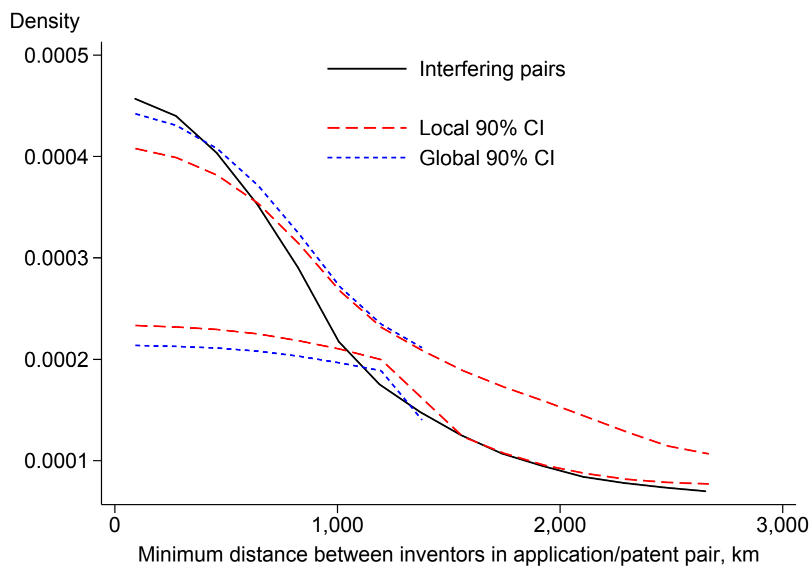


Figure 2: Distribution of geographic distances vs. citation-linked controls

This graph compares the distribution of geographic distances between pairs of interfering inventors to the distribution for control pairs that include one interfering patent and one patent cited by the interfering patent. The black line shows the estimated kernel density function for all interfering pairs for which we were able to find suitable controls. The dotted red lines show the local confidence bands of the kernel density for non-interfering control pairs. The dotted blue lines show the global confidence bands of the kernel density for non-interfering control pairs. For each interfering and control pair we used the minimum pairwise distance among co-inventors. Control patents are restricted to backwards citations from one of the interfering patents.

perform matching-rate regressions, following Henderson, Jaffe, and Trajtenberg (1993). By using regression, we can condition the localization of interfering pairs to better understand the added value of using interferences compared with other measures of knowledge spillovers. We can also assess the extent to which observed previous co-inventor ties might account for the role of geography in interference.

We pool our interfering and control inventor pairs into a database of pairs, indexed by i. Then we estimate

$$1(\text{Distance}_{i[g]} < \Delta) = \beta_1 1(\text{Interfering}_i) + \mathbf{X}_i\beta_x + \mu_g + \epsilon_i \qquad (1)$$

where $1(\text{Distance}_{i[g]} < \Delta)$ is an indicator for whether the minimum distance between interfering inventor teams are within $\Delta = 1.6$, 80, or 161 kilometers (1, 50, or 100 miles) of each other, $1(\text{Interfering}_i)$ is an indicator denoting whether a pair involves interfering patents, and X_i is a vector of attributes of the pair. A group fixed effect μ_g ensures that β_1 is identified based on comparisons *within* a pair group defined by an interfering inventor pair and all associated control pairs. The coefficient β_1 thus estimates how much more likely an interfering inventor pair will "match" in location compared with the average of associated control pairs. We cluster standard errors at the pair-group level.

Table 10 shows estimates from six univariate fixed-effects regressions that vary the set of control patents and the geographic matching threshold. We use three thresholds for defining a geographic match between pairs of inventor teams. The broadest is an indicator equal to 1 when the minimum distance between the locations of residence reported by an interfering or control pair is within 161 km (100 miles), roughly the size of a very large metropolitan area. The narrowest is an indicator equal to 1 when the minimum distance between an inventor pair is within 1.6 km (1 mile). For example, a pair of interfering or control inventors who report living in neighboring municipalities would qualify as a geographic match under the broad definition but not the narrow one. An intermediate geographic matching definition is 80 km (50 miles).

The estimates confirm the patterns seen in Table 9.[13] For all combinations of control patents and distance matches, interfering pairs are at least 1.4–2 times more likely compared with control pairs to match on reported location. The differences are in nearly all cases precisely estimated. For example, 6.3 percent of all interfering and 6-digit control pairs had

[13]The reported geographic matching rates differ slightly in that Table 9 reports estimates based on simulated draws of the pool of eligible control patents separately for interfering and control pairs while Table 10 reports simple means of the entire pool of interfering and control pairs.

inventors reporting locations within 80 km of each other, and interfering inventor pairs were 4.6 percentage points more likely to match at this distance.[14]

	(1) 3-digit controls		(2) 6-digit controls	
	μ [σ]	$\hat{\beta}_1$	μ [σ]	$\hat{\beta}_1$
Minimum distance between pair is within ... km				
161km	0.134	0.0806^c	0.077	0.0549^c
(100mi)	[0.340]	(0.0123)	[0.266]	(0.0123)
80km	0.043	0.0716^c	0.063	0.0462^c
(50mi)	[0.203]	(0.0114)	[0.244]	(0.0115)
1.6km	0.009	0.0180^b	0.014	0.0062
(1mi)	[0.092]	(0.0060)	[0.118]	(0.0060)
Pairs		5,712,342		604,828
Pair-groups		831		821

Table 10: Effect of interference on geographic matching between paired inventors
Each cell is an estimate from a separate fixed-effects regression. The dependent variable is an indicator that equals 1 when the minimum distance between a pair of inventor teams is less than the distance indicated by the row heading. The explanatory variable is an indicator that equals 1 when the pair is an interfering pair and 0 when the pair is a control pair. The sample is interfering pairs of inventors from priority decisions and concessions only, along with their associated control pairs. The set of control pairs is determined by control patents matched on 3- or 6-digit technology class, as indicated by the column heading. Columns labeled μ [σ] show sample means and standard deviations of the geographic matching rates for the pooled sample of interfering and 3- or 6-digit control pairs. Robust standard errors, clustered on pair-group, in parentheses. ^{a}—$p < 0.10$, ^{b}—$p < 0.05$, ^{c}—$p < 0.01$.

In Table 11, we show geographic matching results conditioned on other factors. Column (1) repeats the estimates from Table 10 for matching within 80 km and 6-digit control pairs. Column (2) includes controls for the number of USPTO technology classifications and subclassifications shared by the interfering or control pair of inventions. We also include a control for the number of shared (backwards) patent citations. Even conditioned on textual measures of invention similarity, interfering inventor pairs are still more likely to match geographic location compared with control pairs. This estimate echoes the distance-based results presented in Figure 2. Thus, it also suggests that interferences may better capture spillovers of tacit knowledge compared with commonly used measures use to track knowledge

[14]We also use the matching-rate regression to examine how the degree of geographic localization varies by broad technology categories. Results are shown in Appendix Table 1. Our estimates are imprecise, but suggest that the geographic localization of interfering pairs is not limited to a single sector.

flows.

	μ [σ]	(1)	(2)	(3)	(4)
1(Interfering$_i$)	0.003	0.0462^c	0.0327^b	0.0475^c	0.0378^b
	[0.052]	(0.0115)	(0.0119)	(0.0113)	(0.0116)
Number of shared	1.14		0.000888		-0.0000470
technology classes	[0.84]		(0.00199)		(0.00204)
Number of shared	0.79		0.00782^c		0.00551^c
subclasses	[1.08]		(0.00169)		(0.00157)
Number of shared	0.056		0.00477^a		0.00371^a
backwards citations	[1.68]		(0.00206)		(0.00168)
1(Network$_i$)	0.33			0.232^c	0.229^c
	[0.47]			(0.0169)	(0.0168)
1(Network$_i$)	3.54			-0.0169^c	-0.0167^c
$\times$ Network distance	[5.68]			(0.00128)	(0.00128)
Pairs		604,828	604,828	604,828	604,828
Pair-groups		821	821	821	821

Table 11: Effect of interference on geographic matching conditioned on textual links and co-inventor ties

Each column is an estimate from a separate fixed-effects regression. The dependent variable is an indicator that equals 1 when the minimum distance between a pair of inventor teams is less than 80km (50mi). 1(Interfering$_i$) is an indicator that equals 1 when the pair is an interfering pair and 0 when the pair is a control pair. 1(Network$_i$) is an indicator that equals 1 when the pair is connected in the backwards-looking co-inventor network. Network distance is the number of edges or vertices between paired inventors in the co-inventor network. The sample is interfering pairs of inventors from priority decisions and concessions only, along with their associated control pairs. The set of control pairs is determined by control patents matched on 3- or 6-digit technology class, as indicated by the column heading. Column labeled μ [σ] shows sample means and standard deviations of the explanatory variables. Robust standard errors, clustered on pair-group, in parentheses. ^{a}—$p < 0.10$, ^{b}—$p < 0.05$, ^{c}—$p < 0.01$.

Next, we examine the degree to which social network ties drive the geographic localization of interferences. To shed light on this mechanism, we follow Breschi and Lissoni (2009), and consider past co-inventorship as a proxy for social ties. These authors found that controlling for this measure of social network ties greatly reduced estimates of the geographic localization of citations. More generally, Head, Li, and Minondo (2015) find that controlling for measures of the network ties of mathematicians halves the estimated impact of geographic distance on citations.

We define a social network with each inventor represented as a node and connections or edges between any inventors that have been co-inventors on an issued patent. We use name disambiguation available in the Lai et al. (2013) database to identify unique inventors. (Their algorithm uses not only name similarity but also inventor location, assignee, and technological class information).

The network distance between two inventors is the minimum path distance between them in the network—the number of edges along the shortest path from one inventor node to the other. This network distance is conditioned on co-inventor links up to 5 years *before* the earliest application date in interference. For a pair of patents A and B, we assign the shortest social network distance between any inventor on patent A and any inventor on patent B.

Columns (3) and (4) of Table 11 condition on co-inventor ties. There are two included controls: One, an indicator for whether an inventor pair is linked in our constructed backwards-looking co-inventor networks. Two, an interaction between this indicator and the number of edges along the shortest path from one inventor node to the other.

Previous co-inventor ties predict geographic matching, as evidenced by the positive and precisely estimated coefficient on the indicator for a network connection. Inventors connected through co-inventor ties are 23 percent more likely to live within 80km. This relationship attenuates with network distance by 1.7 percent for each additional edge or node in the co-inventor network. Interestingly, conditioning on observed co-inventor ties does not change the estimated effect of interference on geographic matching. In contrast to Breschi and Lissoni (2005), we find little evidence that the localization of interfering inventors is mediated via social ties, at least as proxied by previous co-inventorship[15].

Of course, social and professional ties and location are endogenous choices, so our estimates cannot be interpreted as causal effects. It is also likely that knowledge flows among *unobserved* ties between inventors. Therefore, we view this specification as accounting for the role of co-inventor ties in mediating localized knowledge spillovers. In contrast to Breschi and Lissoni (2005) and Head, Li, and Minondo (2015), we find only modest evidence that the localization of interfering inventors is mediated via social ties, at least as proxied by previous co-inventorship. As previous results rely on citations as measures of spillovers, our results emphasize a continued role for geographic distance in knowledge flows.

[15] Appendix Figures 1 and 2 show evidence that interfering patents are more likely than control patents to be connected by social network ties, and some evidence that they are likely to be closer in social network space. However, the regression results in Table 11 suggest that this social network localization is independent of geographic localization.

5 Conclusions

We present new evidence of localized knowledge spillovers using a novel database of patent interferences—instances of simultaneous, identical invention by multiple, independent parties. By evidence of common, identical invention, interfering inventors share common knowledge inputs. Interfering inventor pairs show significant geographic localization compared with the counterfactual of inventor pairs sharing similar invention dates and technology classification. Thus, our results provide verification of the existence of localized knowledge spillovers. We also overcome challenges identified by the literature using patent citations that "one-half of all citations do not correspond to any spillover" (Jaffe, Trajtenberg, and Fogarty, 2000) and "an enormous number of spillovers [occur] with no citations" (Jaffe, Trajtenberg, and Henderson, 1993). Interfering inventor pairs are even more localized compared with cited-citing inventor pairs, consistent with Jaffe, Trajtenberg, and Henderson's (1993) conjecture that citations are a lower bound on the strength of localized knowledge spillovers.

Our results suggest that, in contrast to conventional wisdom about "the death of distance," geographic distance continues to matter, especially for flows of tacit, or difficult to codify, forms of knowledge. These are the types of knowledge flows where the lack of a "paper trail" has hampered the availability of evidence. Interferences therefore provide a unique and useful window into localized knowledge spillovers. In future work, it would be useful to leverage the potential of interferences to measure shared knowledge inputs to investigate other features of the invention process.

6 Works cited

Arzaghi, Mohammad, and J. Vernon Henderson (2008). "Networking off Madison Avenue," *The Review of Economic Studies*, 75 (4): 1011–1038.

Audretsch, D.B. and M.P. Feldman (2004). "Knowledge Spillovers and the Geography of Innovation," *Handbook of Regional and Urban Economics*, vol. 4, 2713–2739.

Bikard, Michaël (2012). "Simultaneous Discoveries as a Research Tool: Method and Promise," working paper.

Breschi, Stefano and Francesco Lissoni (2005). "'Cross-Firm' Inventors and Social Networks: Localized Knowledge Spillovers Revisited," *Annals d'Économie et de Statistique* 79–80: 189–209.

Buzard, Kristy, Gerald A. Carlino, Robert M. Hunt, Jake K. Carr, and Tony E. Smith (2017). "The Agglomeration of American R&D Labs," *Journal of Urban Economics* 101: 14–26.

Calvert, Ian A. (1980). "An Overview of Interference Practice," *Journal of the Patent Office Society* 62: 290–307.

Calvert, Ian A. and Michael Sofocleous (1982). "Three Years of Interference Statistics," *Journal of the Patent Office Society* 64 (12): 699–707.

Calvert, Ian A. and Michael Sofocleous (1986). "Interference Statistics for Fiscal Years 1983 to 1985," *Journal of the Patent and Trademark Office Society* 68 (8): 385–393.

Calvert, Ian A. and Michael Sofocleous (1989). "Interference Statistics for Fiscal Years 1986 to 1988," *Journal of the Patent and Trademark Office Society* 71 : 399-410.

Calvert, Ian A. and Michael Sofocleous (1992). "Interference Statistics for Fiscal Years 1989 to 1991," *Journal of the Patent and Trademark Office Society* 74 : 822–826.

Calvert, Ian A. and Michael Sofocleous (1995). "Interference Statistics for Fiscal Years 1992 to 1994," *Journal of the Patent and Trademark Office Society* 77 : 417–422.

Carlino, Gerald A., Jake K. Carr, Robert M. Hunt, and Tony E. Smith (2017). "The Agglomeration of American R&D Labs," *Journal of Urban Economics* 101: 14–26.

Carlino, Gerald A., Satyajit Chatterjee, and Robert M. Hunt (2007). "Urban Density and the Rate of Invention," *Journal of Urban Economics* 61: 389–419.

Carlino, Gerald A. and William R. Kerr (2015). "Agglomeration and Innovation," in *Handbook of Regional and Urban Economics* vol. 5, Gilles Duranton, J. Vernon Henderson, William C. Strange, eds., Elsevier, 349–404.

Chatterji, A., E. Glaeser, and W.R. Kerr (2014). "Clusters of entrepreneurship and innovation," in *Innovation Policy and the Economy*, vol. 14, J. Lerner, S. Stern, eds., University of Chicago Press, 129–166.

Cohen, Linda R. and Jun Ishii (2006). "Competition, Innovation, and Racing for Priority at the U.S. Patent and Trademark Office," working paper.

Combes, Pierre-Philippe and Laurent Gobillon (2015). "The Empirics of Agglomeration Economies," in *Handbook of Regional and Urban Economics*, vol. 5, Gilles Duranton, J. Vernon Henderson, William C. Strange, eds., Elsevier, 247–348.

De Simone, Daniel V., James B. Gambrell, and Charles F. Gareau (1963). "Characteristics of Interference Practice," *Journal of the Patent Office Society* 45: 503–591.

Duranton, Gilles and Henry Overman (2005). "Testing for Localization Using Micro-Geographic Data," *Review of Economic Studies* 72: 1077–1106.

Duranton, Gilles and Diego Puga (2004). "Micro-foundations of Urban Agglomeration Economies," in J. Vernon Henderson and Jacques-François Thisse (eds.), *Handbook of Regional and Urban Economics*. Amsterdam: Elsevier, vol. 4, chap. 48, 2063–2117.

Hall, Bronwyn H., Adam B. Jaffe, and Manuel Trajtenberg (2001). "The NBER Patent Citation Data File: Lessons, Insights, and Methodological Tools," NBER Working Paper 8498.

Head, Keith, Yao Amber Li, and Asier Minondo (2015). "Geography, Ties, and Knowledge Flows: Evidence from Citations in Mathematics," working paper.

Feldman, M.P. and D.B. Audretsch (1999). "Innovation in Cities:: Science-based Diversity, Specialization and Localized Competition," *European Economic Review*, 43(2), 409–429.

Gladwell, Malcolm (2008). "In the Air," *The New Yorker*, May 12.

Jaffe, Adam B., Manuel Trajtenberg, and Michael S. Fogarty (2000). “Knowledge Spillovers and Patent Citations: Evidence from a Survey of Inventors,” *American Economic Review* 90(2): 215–218.

Jaffe, Adam B., Manuel Trajtenberg, and Rebecca Henderson (1993). “Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations,” *Quarterly Journal of Economics* 108 (3): 577–598.

Kingston, William (2001). “Light on Simultaneous Invention from U.S. Patent Office ‘Interference’ Records,” *World Patent Information* 26: 209–220.

Kleid, Dennis G (2002). “Scientist and Patent Agent at Genentech,” an oral history conducted in 2001 and 2002 by Sally Smith Hughes for the Regional Oral History Office, The Bancroft Library, University of California, Berkeley.

Krugman, Paul (1991). *Geography and Trade.* Cambridge: MIT Press.

Lai, Ronald, Alexander D’Amour, Amy Yu, Ye Sun, and Lee Fleming, 2013,“Disambiguation and Co-authorship Networks of the U.S. Patent Inventor Database (1975 - 2010)”, `http://hdl.handle.net/1902.1/15705`. The Harvard Dataverse Network [Distributor] V5 [Version].

Lin, Jeffrey (2014). “The Paper Trail of Knowledge Transfers,” Federal Reserve Bank of Philadelphia *Business Review* Q2: 1–6.

Lucas, Robert (1988). “On the Mechanics of Economic Development,” *Journal of Monetary Economics* 22: 3–42.

Marshall, Alfred. *Principles of Economics.* London: Macmillan, 1920.

Merton, Robert K. (1973). *The Sociology of Science: Theoretical and Empirical Investigations.* Chicago: University of Chicago Press.

Murata Yasusada, Ryo Nakajima, Ryosuke Okamoto, and Ryuichi Tamura (2014). “Localized knowledge spillovers and patent citations: A distance-based approach,” *Review of Economics and Statistics* 96:5, 967–985.

Ogburn, William F. and Dorothy Thomas (1922). “Are Inventions Inevitable? A Note on Social Evolution,” *Political Science Quarterly* 37 (1): 83–98.

Pollack, Andrew (2017). "Board Rules on Patents for Process to Edit DNA." *The New York Times*, February 16, page B3.

Rosenthal, Stuart S., and William C. Strange (2004). "Evidence on the Nature and Sources of Agglomeration Economies", in J. Vernon Henderson and Jacques-François Thisse (Eds.), *Handbook of Regional and Urban Economics* (Amsterdam: Elsevier), 2119–2171.

Schmookler, Jacob (1966). *Invention and Economic Growth*, Harvard University Press.

Thompson, Peter (2006). "Patent Citations and the Geography of Knowledge Spillovers: Evidence from Inventor- and Examiner-Added Citations," *Review of Economics and Statistics* 88 (2): 383-388.

Thompson, Peter, and Melanie Fox-Kean (2005). "Patent Citations and the Geography of Knowledge Spillovers: A Reassessment." *American Economic Review* 95 (1): 450–460.

United States Court of Appeals for the Federal Circuit. Decision 99-1604 in interference no. 102416, *Hitzeman et al. v. Rutter et al.* March 21, 2001.

United States Patent and Trademark Office. "e-FOIA Reading Room," http://e-foia.uspto.gov/Foia/PTABReadingRoom.jsp, accessed November 3, 2014.

United States Patent and Trademark Office. "Efiling for Patent Trial and Appeal Board," https://acts.uspto.gov/ifiling/PublicView.jsp, accessed May 6, 2015.

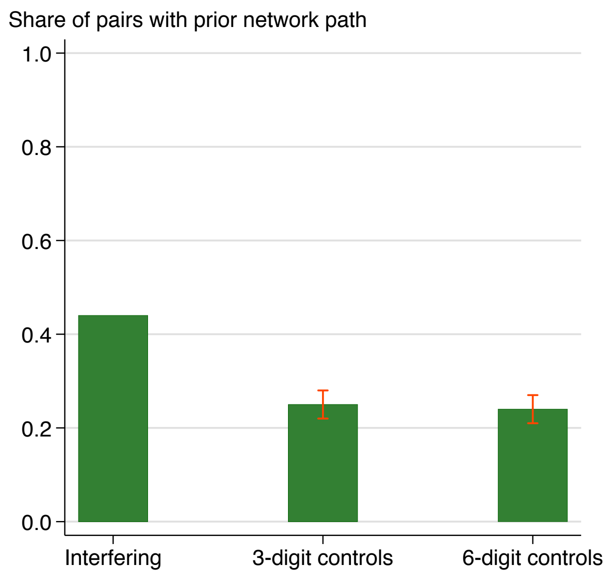
Weitzman, Martin L. (1998). "Recombinant Growth," *Quarterly Journal of Economics* 113 (2): 331–360.

Appendix

	3-digit	6-digit
Chemical	0.0642^b	0.0404^a
	(0.0201)	(0.0199)
Computers and communications	0.0502	0.0539
	(0.0543)	(0.0528)
Drugs and medical	-0.00655	-0.000793
	(0.0268)	(0.0270)
Electrical and electronic	0.00977	0.00500
	(0.0449)	(0.0447)
Mechanical	0.0126	-0.0128
	(0.0438)	(0.0437)
Others	0.0701	0.0467
	(0.0531)	(0.0522)
Pairs	5,709,043	604,692
Pair-groups	821	811

Appendix Table 1: Effect of interference on geographic matching by technological category

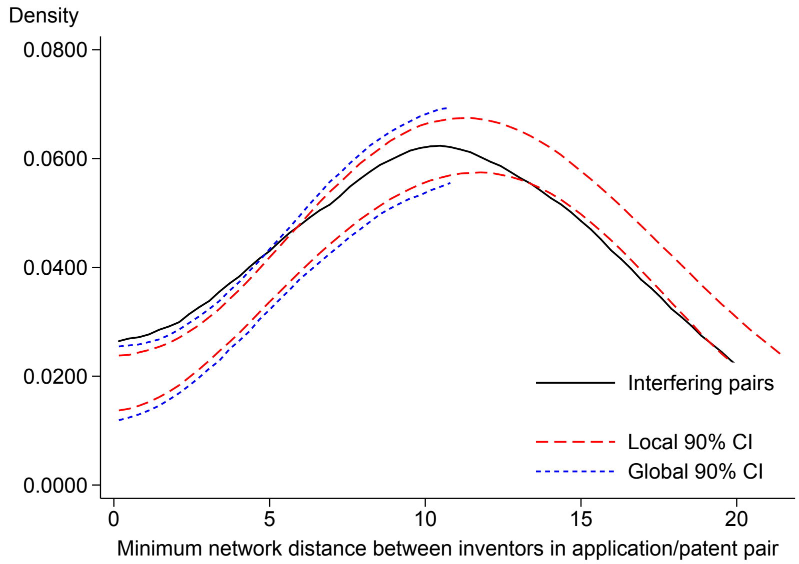
Each column is an estimate from a separate fixed-effects regression. The dependent variable is an indicator that equals 1 when the minimum distance between a pair of inventor teams is less than 80km (50mi). Each row shows the coefficient from an interaction between a dummy for the listed technology class with a dummy for interfering pairs. The technological categories come from Hall, Jaffe and Trajtenberg (2001). The sample is interfering pairs of inventors from priority decisions and concessions only, along with their associated control pairs. The set of control pairs is determined by control patents matched on 3- or 6-digit technology class, as indicated by the column heading. Robust standard errors, clustered on pair-group, in parentheses. ^{a}—$p < 0.10$, ^{b}—$p < 0.05$, ^{c}—$p < 0.01$.



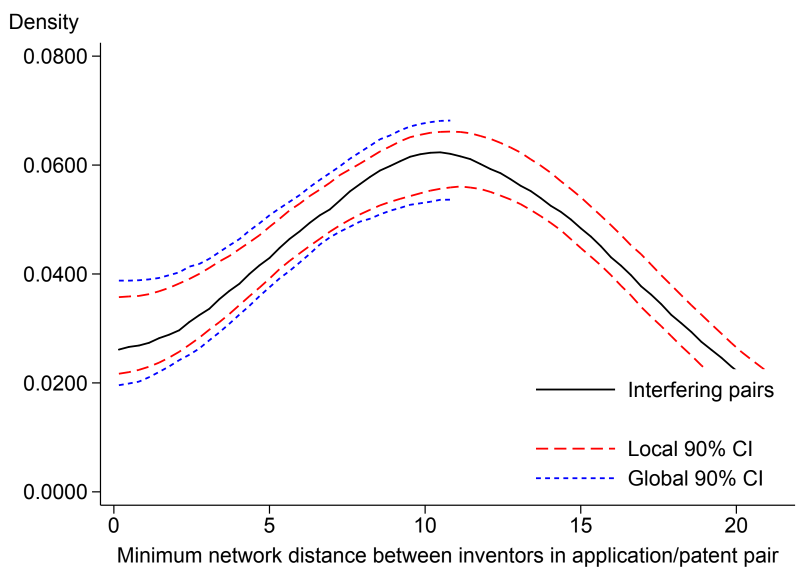
Appendix Figure 1: Interfering inventors are more likely to be connected by previous co-inventor ties

Priority decisions only. Simulated 90 percent confidence intervals shown.

Panel A. 3-digit controls



Panel B. 6-digit controls



Appendix Figure 2: Density of co-inventor network distance between inventor pairs

Notes: This graph compares the distribution of co-inventor network distances between pairs of interfering inventors to the distribution for similar, non-interfering control pairs. The black line shows the estimated kernel density function for all interfering pairs for which we were able to find suitable controls. The dotted red lines show the local confidence bands of the kernel density for non-interfering control pairs. The dotted blue lines show the global confidence bands of the kernel density for non-interfering control pairs. For each interfering and control pair we define network distance as the length of the shortest distance path along our constructed inventor social network between any two co-inventors. Control patents are restricted to share a 6-digit technology class with both interfering patents. The sample is priority decisions only.