PRELIMINARY DRAFT

Measuring the Diffusion of Innovation: A Reassessment of Knowledge Spillovers Using Machine Learning

Fiona Sijie Feng (NYU Stern)

October 11, 2018

Click here for latest version

Abstract

The ideas of new invention are captured by the text of a patent. Empirical measures of knowledge flows have relied on citations, which reflect the influence of other patents on an invention but not external sources of knowledge. I use unsupervised machine learning methods to convert patent abstracts (which are descriptions of the invention) to numerical vector space representations. Knowledge spillovers are measured using the similarity across patent text vectors. I find that geographic localization effects are insignificant to modest: prior to controlling technological proximity, within technology field patents from the same city are about 0.03-0.08 standard deviations more similar than patents from different cities. Including further technology controls reduces estimates from insignificant to 0.04. By contrast, citations based measures find that local patents have 0.24-0.30 standard deviations more citations from the same city compared to a non-local control. I also find evidence that differentiation may play a minor role in determining innovation for technologically proximate rivals seeking to broaden the scope of their patents, and that this motivation is more pronounced for local rivals. These findings indicate that geographic localization of knowledge spillovers (i) may not be a strong driver of agglomeration, as suggested by standard citations-based analyses; (ii) affects local innovation in a nuanced manner.

1. Introduction

Knowledge spillovers are widely believed to be one of the key contributors to the phenomena of agglomeration. As cities and regional economies occupy a growing significance as populations rapidly urbanise worldwide, it is imperative that the underlying factors responsible for the dynamism of cities is understood. Innovation has become increasingly concentrated in major urban centers: accounting for 40% of new patents in 1975-85 to 46% of new patents 2005-15. One explanation for why innovation is an urban phenomena is that knowledge spillovers are geographically localized. This means that local innovation generates greater positive externalities for local firms over non-local firms. A prominent literature of measuring knowledge spillovers has emerged from Jaffe et al. (1993) (henceforth JTH) that uses patent citations to study the "paper trail" left by the diffusion of innovative knowledge. The general consensus of this literature is that there are sizable geographic localization effects for knowledge spillovers.

This paper uses methodology from unsupervised machine learning to analyze the ideas in patent abstracts, which summarize the invention. I use these methods to make three key contributions: first, I derive numeric vector space representations of innovative ideas using the Document Vectors (Doc2Vec) algorithm. Second, instead of using citations, I proxy for knowledge spillovers by analyzing the cosine similarity across patent vectors. Third, I derive novel measures of technological proximity by using average similarity across patent groups (i) within the same USPTO primary class; (ii) within the same primary class and same metropolitan statistical area (MSA).

I find that the evidence for local knowledge spillovers is mixed. Without technological proximity controls, local patents are 0.03-0.08 S.D.s more similar to other local patents compared to non-local patents; with such controls, the localization estimate ranges from insignificant to 0.04 S.D.s. Time trends are roughly similar across a range of specifications: localization is lowest in 1975-85, higher in 1985-2005, then falling 2005-15. In comparison, the citations-based benchmark finds that patents receive 0.24-0.30 SD more local citations compared to non-local patents which rises over the timeframe 1975-2005.

For patents within the same primary class, I find that when technology proximity across locations is high: (i) average similarity *declines*; (i) average similarity is lower for patents from the same location compared to patents from different locations. If we consider patents from technologically proximate locations to be potentially close rivals, both these facts indicate that some *differentiation* from rivals occurs and that such effects are likely more pronounced for local rivals. Differentiation from rivals in innovation may be driven by two considerations: (i) inventors seeking to broaden the scope of their patents; (ii) inventors seeking to avert potential infringement claims. I find some evidence that local differentiation is most prevalent in technology fields (such as pharmaceutical drugs) where reliance on patent protection is strongest.

One concern is that rather than geographic match being a weak determinant of knowledge relatedness across patents, similarity itself is a noisy proxy. I address this by seeing if similarity is able to capture strong determinants of knowledge relatedness by matching in other dimensions. For patent pairs within the same industry, I find that pairs from the same primary class (a narrower technology group) increases similarity by 0.4-0.44 S.D.s; sharing an inventor (who has changed firms) increases similarity by 1.27-1.52 S.D.s. Thus, since similarity is able to uncover variables that do have a strong effect on knowledge relatedness and the weakness of geographic proximity cannot be attributable to noise.

These findings imply that current methods of measuring knowledge spillovers may overestimate localization by focusing only on citations. For a given patent, any number of prior patents may be relevant enough to serve as a "citation." Inventors and firms draw from their own prior knowledge stock of patents and inventors to form a list of citations, but this list is by no means exhaustive of the relevant knowledge flows for the new invention. Packalen and Battacharya (2015) and Rosenberg (1982) emphasize the fact that external knowledge sources also play a role in the innovation process. If such knowledge flows are non-local in nature, then this leads to less localization across patents, even if citation networks are local.

I examine the influence of external knowledge flows by looking at the first patents to use new technological terms. I find that while similarity captures the overlap in the ideas expressed by inventions appropriating the same new technology, these patents have almost no backward citations in common. Citations would suggest that these patents arose from completely unrelated knowledge sources. However, analysing the text of patents indicates that patents with no common citation link may still produce similar innovations through the influence of external knowledge flows. For example, “adenorivus” patents used in gene therapy treatment first appeared in patent applications in 1993. While these patents have an average similarity of 0.26, they share an average of 0.03 or 0.0% of backward citations in common. This points to external knowledge flows playing a significant role in accounting for the difference between citation and similarity measures of knowledge spillovers.

Alongside Jaffe et al. (1993), the prior literature using has found significant localization in a variety of contexts: Murata et al. (2014) and Buzard et al. (2016) using spatial distance measures; Almeida and Kogut (1999), Agrawal et al. (2006), and Azoulay et al. (2011) using geographic mobility of inventors; Belenzon and Schankerman (2013) with university patents and scientific publications. Only Thompson and Fox-Kean (2005) find that localization estimates are insignificant between extremely technologically proximate patents.

I explore what might account for the difference in the localization estimates using citations and similarity. One possibility, as discussed in Jaffe et al. (2000), is that citations are weak indications of knowledge flows. There are three main sources of interference: (i) the addition of citations from patent examiners as discussed in Alcacer et al. (2009); (ii) strategic considerations to add irrelevant citations to block potential infringement suits and to omit relevant citations to broaden the patent scope (Lampe (2012), Roach and Cohen (2013)); (iii) the influence of lawyers on applicant's citations (Moser et al. (2017), Wagner et al. (2014)).

I find evidence that strategic considerations may overstate the localization of citations: local patents represent a greater proportion of less relevant citations, while non-local patents are drastically under-cited when they are highly relevant. This is consistent with the expectation that the possibility of discovering infringement is higher when firms are located closely (Lin et al. (2014)). Furthermore, I examine changes in knowledge flows using citations and similarity when inventors switch firms. I find that inventors consistently cite their own previous inventions less after shifting firms, particular for highly similar prior patents. This points to the presence of strategic omissions as inventors cannot be expected to have "reasonable ignorance" of their own prior inventions.

Additionally, I explore the change in knowledge flows as measured by citations and similarity when the inventor changes firms. I find that the proportion of common backward citations made by the same inventor drastically declines after changing firms even within the same city: from sharing 13% of backward citations prior to the move to 6% after. This would represent a 54% change in knowledge flows used by the inventor. The similarity to the inventor's own prior patents declines more modestly: from 0.29 to 0.25, or a 14% change. This suggests that firm specific factors, such as the firm's knowledge of prior patents, and their choice of lawyers, determine a large proportion of citations. I find evidence that lawyers play a large role in determining a firm's citations: patents from different firms sharing the same lawyer has 1.2-4.3% higher backward citations in common.

When inventors move across cities, I find that their forward citations from the new city rises from

3% to almost 8%. However, I find limited evidence that the knowledge flow from the mobile inventor to the newly citing firm is significant, as the patents that cite the mobile inventor fall well within the existing innovation agenda of the firm. In line with my main results, this points to the size of local knowledge spillovers being relatively small.

Roadmap The paper proceeds as follows. Section 2 discusses the data sources, and outlines the NLP methodology. Section 3 assesses citation patterns when examined in conjunction with cross-patent similarity. I explore implications for citations using inventor mobility in section §4. Section 5 presents the estimation models and results for regressions. Section 6 examines the influence of external sources of knowledge through new technology terms appearing in the patent corpus. I discuss policy implications in 7. Further tables and graphs can be found in the appendix.

2. Data and Methodology

2.1. Data Sources

Patent data is taken from PatentsView on all utility patents granted 1976-2016, containing data both on inventors (including unique identifiers and location) and patents (assignee, application date, grant date, primary class and subclass). Bibliographic text data is taken from the USPTO Bulk Data Products, which has all patent bibliographic text from 1976 to end of 2015. Patent abstracts are taken to be representative of the knowledge contained in patents, as they are a summary of the invention. Following JTH, the patent's location is determined as the MSA where the highest proportion of inventors are located.

Patent technology fields Each patent is assigned three technological *fields*, with each field being nested in the previous. At the broadest level, an NAICS-based industry classification is given using the USPC to NAICS concordance crosswalk, which delegates each patent to a NAICS category according to its USPTO 3-digit primary classification. Additionally, many patents are also assigned a primary *sub*class.[1] Primary subclasses are nested in primary classes, which are in turn nested in a NAICS industry label. There are over 150,000 subclass labels; 450 class labels, and 33 NAICS industry labels.

[1] Patents may also include other discretionary classifications, which are not used in my data.

2.2. Patent Abstracts to Vector Space Representations

Using patent abstract texts, I use procedures standard in the NLP literature to clean and convert text to vector representations (see section A.1 for details). I focus on Document Vectors generated by Doc2Vec, introduced by Le and Mikolov (2014) and also apply Latent Dirichlet Allocation to validate the findings for Document Vectors. Since the results for the two measures are aligned and Doc2Vec performs better in a range of validation exercises, I direct my discussion using patent vectors from Doc2Vecs.

Natural Language Processing and Patents Literature The use of natural language processing in business and social science fields is still in its nascency. Most notably, Kaplan and Vakili (2015) fit a Latent Dirichlet Allocation model on patent text to determine breakthrough innovation. They find that a topic-originating or breakthrough patent receives approximately 1.4 times more citations than the average patent. Bergeaud et al. (2017) also use relevant keywords found in patent abstracts to construct a semantic network classification system to map the technological taxonomy of patents. Bryan et al. (2018) use a machine learning approach to identify in-text citations within patent text, which they propose as a better measure of direct knowledge flows. Without using machine learning methods, Packalen and Battacharya (2015) use the appearance of new terms to examine how adoption of new technology varies by city size.

Unsupervised Machine Learning Methods: Document Vectors

The Doc2Vec algorithm was introduced by Le and Mikolov (2014) as a means to meaningfully summarize text contained within documents. It is a straightforward extension of the word2vec model of Mikolov et al. (2013b,a), which was developed to represent words meaningfully in a vector space (provide "word embeddings"). The objective of word2vec is to situate words that have similar meanings close to one another. Similarly, Doc2Vec has the objective of situating similar *documents* close to one another by placing document vectors (DocVec) close to each other in vector space. To do this, the algorithm uses the "context" around each term in the document to derive a vector representation that maximizes the probability its the appearance. (See A.1.1 for more details on the algorithm; figure A.2 illustrates diagramatically the inputs and outputs of the algorithm) I implement the algorithm using the `gensim` package in Python (Řehůřek and Sojka (2010)).

For example, for the sentence “Provides for unattended file transfers”, the central word “unattended” has the context [“Provides”, “for”, “file”, “transfers”]. Before the algorithm is implemented, common words or stop words such as “for” are removed and each word is stemmed to the root. “Provides” and “transfers” become “provid” and “transfer.” The document identifier, in this case the patent number “US7502754,” is treated as a context word for ever word in the patent. Thus, the context for “unattended” would become: [“provid”, “file”, “transfer”, “US7502754”].

Every word and document is assigned a vector of dimension $N = 100$.[2] The vectors are optimized using a neural network which maximises the log probability of the appearance of each central word. The resulting vector places words that arise in similar contexts close to each other, and documents that contain similar words close to each other.

2.3. Measuring Knowledge Spillovers: Cross Patent Similarity

Cosine similarity [3] has been used to measure technological proximity in Jaffe (1989) and Bloom et al. (2013), as well as being standard in the NLP literature (Mihalcea et al. (2006)). The prior literature used vectorizations of patent classes listed for each patent, which had the issues of being of varying lengths with unassigned weights for each class. The primary advantage of NLP patent vector outputs is that they are *jointly* determined, and position each patent vector *relative* to all other patents within the corpus. Thus, cross-patent comparisons using NLP vector outputs are much more internally consistent than using vectorizations of patent class selections.

For two patents, i and j, the cosine similarity between them is:

$$sim(i,j) = \frac{PV_i \cdot PV_j}{\|PV_i\| \|PV_j\|} \tag{2.1}$$

Where PV_i is the patent vector representation of i. This is preferred to Euclidean distance as it is factors in the “size” of the vector; a Euclidean distance measure would assign positive distance to two vectors that contained the exact same words, but of different quantities. Cosine similarity normalises all measures to be in the range $[-1,1]$.

[2]This is a rule-of-thumb in the literature, according to Lin et al. (2015)

[3]Other measures, such as Hellinger distance, were also used but found to be very highly correlated with cosine similarity.

Analogous Citations Measure

Since similarity may be interpreted as the degree of common knowledge between two patents, an analogous measure using citations would be the number or proportion of common backward citations between two patents:

$$ncc(i,j) = |\{citations_i\} \cap \{citations_j\}| \tag{2.2}$$

$$pcc(i,j) = \frac{|\{citations_i\} \cap \{citations_j\}|}{|\{citations_i\}|} \tag{2.3}$$

Where $ncc(i,j)$ represents the number of common backward citations between patents i,j and $pcc(i,j)$ the proportion of backward citations of $i$that were also made by j. For example, if patent i cites $\{A,B,C,D\}$, and j cites $\{D,E\}$, $ncc(i,j) = 1$ and $pcc(i,j) = 0.25$. In each case, self-citations are removed first. These variables can be thought of as measuring the degree of similitude in the knowledge sources of the two patents using a citations-based approach.

Technological Field Proximity

Since each patent is assigned technology field labels in the form of primary classes, technological field proximity between two primary classes can be measured using the average similarity of a sample of patents in each primary class. For each year t, I take a sample of up to 1000 patents in each primary class pair pc_i, pc_j that were granted in the previous 5 years. I then calculate the mean of the pairwise similarities between all such pairs. Thus:

$$sim(pc_i, pc_j)_t = mean\left(\{sim(i,j) | i \in pc_i, j \in pc_j\}_{t-5,t}\right) \tag{2.4}$$

Self-similarity where $sim(i,j) = 1$ are removed. Similarly, the technological proximity for primary classes at a specific MSA (primary class-MSA) can be measured as:

$$sim(pc_{i,MSA_i}, pc_{j,MSA_j})_t = mean\left(\{sim(i,j) | i \in pc_i, i \in MSA_i, j \in pc_j, j \in MSA_j\}_{t-5,t}\right) \tag{2.5}$$

Here, patent i is granted within years $t-5$ to t, to primary class pc_i and located in MSA_i. Intuitively, this represents the *expected* similarity between two patents if only their technology field was known. Cross field similarity are analogous to the technological proximity measures of Jaffe (1986); Bloom et al. (2013). Both papers, alongside other citations-based methods of measuring technological

Year Group	NAICS Match	S.D.	Primclass Match	S.D.	Inventor Match	S.D.	Year Match	S.D.
1975-85	0.126	0.137	0.187	0.145	0.301	0.148	0.126	0.138
1985-95	0.124	0.135	0.186	0.145	0.320	0.163	0.124	0.135
1995-05	0.129	0.134	0.196	0.147	0.312	0.158	0.129	0.134
2005-15	0.141	0.136	0.200	0.146	0.310	0.170	0.141	0.136

Table 2.1: Average DocVecs Similarity Within Groups

proximity, rely on the vectorization of PTO classes. These methods may lead to inconsistent results as each patent may have any number of non-primary classifications. The standard procedure has been to normalize or weight each of the classes listed, which discretizes the vector space and leads to discontinuities in the proximity measures.[4]

Validating Similarity Measures

Prior expectations about patent similarity can be used to validate the vectors generated by the Doc2Vec algorithm. In table 2.1, the baseline group average is the average pairwise similarity for patent pairs from within the same NAICS industry granted within 5 years of one another. We should expect that, on average, similarity between patent pairs of the same primary class should be *higher* than pairs within the same NAICS industry, since industry represents a broader definition of technology field. Table 2.1 shows that patents within the same primary class have average similarity around 1.5 times that of patents just within the same NAICS industry. Patent pairs sharing an inventor have 2.5 times the similarity of the baseline group. On the other hand, we should also expect that patent pairs from the same grant year should not have average similarity higher than the baseline, since the time difference between 0 years and 1-5 years is not large enough to have a significant impact on technological difference. Table 2.1 shows there is virtually no difference between average similarity of patents granted in the same year and the baseline. In general, variance is higher in smaller samples such as patent pairs with matching inventors. Since DocVecs captures trends in similarity that matches prior expectations, it is unlikely that results are being driven by noise in the vectors generated by the algorithm.

[4] A patent with one class would be represented by a vector with 1 in the class column and 0 elsewhere; two classes 0.5 in each class column and elsewhere; and so on.

What does patent similarity measure?

Similarity should not be taken as a replacement for citation measures. If we consider the innovation process, it requires a number of knowledge flow inputs:

1. Citations of other patents discovered through inventor networks. (observed)
2. Citations of non-commercial scientific publications discovered through scientific research (observed)
3. Outside sources of knowledge such as workplace training, non-academic technical and trade publications, professional conferences (unobserved)

The output of the innovation process is the invention, as described by the text of the patent. This invention would reflect the knowledge inputs, including what may be unobserved external sources of knowledge. These unobserved knowledge sources may then represent a "wedge" between the knowledge spillovers measured by citation patterns, compared to the similarity measure. High similarity measure between two patents then would indicate that the patents share a greater degree of knowledge influences. Unlike citations, high similarity does not indicate whether or not patent i directly influenced patent j. While this means that similarity is suboptimal as a measure of *direct* knowledge flow, it may be a superior measure for knowledge spillovers, which is an *externality* from the exchange of knowledge flows. I discuss the implications of this interpretation of similarity for identifying localization in 5.3. I discuss the differences in what citations and similarity captures in the next section.

3. Comparing citations and similarity as measures of knowledge spillovers

The validity of citations as a measure of knowledge spillovers are challenged by the existing literature. It has been widely used in practice because, until now, another such measure has not been available. The problems with using patent citations to proxy for knowledge flows have been well documented. The two dominant concerns are: (i) many citations added by external agents (either law firms or patent examiners), which obfuscates the relationship between the patent and citation as a direct knowledge "flow"; (ii) there are strategic reasons for withholding relevant citations. Namely, citing patents that are closely proximate to the invention limits the scope of the patent and thus reduces the value of the intellectual property. These effects can result in substantial measurement error: Alcacer and Gittelman (2006) find that on the average patent, two-thirds of citations are added by the examiner, while Cotropia et al. (2013) find that applicant citations are often ignored by examiners who conduct their own search of prior art. Citations are also strategic in that, according to Jaffe and De Rassenfosse (2017), "although applicants at the USPTO have a duty to disclose what they know, they have no duty to search for prior art and may be better off by remaining ignorant." Inventors seeking to maximise the value of their IP may be inclined to leave out the most relevant citations; Lampe (2012) finds that applicants withhold between 21% to 33% of relevant citations, as determined by the applicant firm's previous citations. Using a survey of lab managers, Roach and Cohen (2013) also find that patent citations are more reflective of a firm's appropriability strategies in ways that are not revealing of "true" knowledge flows.

In addition, concern over the possibility of patent litigation can potentially lead to a rise in spurrious citations. Lerner and Seru (2015) discuss tactics used by practitioners to offset the likelihood of lawsuits: "...patent lawyers sometimes urge weak applicants to employ the "kitchen sink" approach to citations: to cite a wide variety of prior art, burying the relevant stuff under a mountain of irrelevant prior art in the hopes that the time-pressed examiner will not discover it." The combination both the incentive to omit highly relevant citations through either wilfull ignorance or strategy and the inclusion of irrelevant citations further casts doubts on the ability of citations to accurately proxy for knowledge flows.

It is also possible that these incentives drive up the measure of localization using citations. If firms are concerned that the probability of infringement discovery by rivals in the same city are more likely, this may induce a greater rate of citation for local firms. Lin et al. (2014) indeed find that patent interference claims occur more frequently between inventors located close together. The omission of relevant patents located elsewhere may further be defensible through both the defense of plausible ignorance and the lower probability of infringement discovery.

Patent vector similarity may not be subject to the same criticism. Because patent abstracts must be accurate summaries of the invention at hand, this limits the ability of applicants to omit important technological terms in order to hide the relevance of previous knowledge. Legal considerations could still play a role in determining how inventions are described: it is likely that applicants may choose words to distance their inventions from a handful of closely related patents. However, since similarity can be determined for *any* pair of patents, the ability for applicants to internalize their choice of terms relative to the entire patent corpus is limited. On the other hand, applicants have complete choice over their list of relevant prior art, which are difficult to hold accountable to an external criteria of accuracy. The authority of the patent examiner to make additions to citations list is precisely a measure enacted to counteract this problem.

3.1. Effect of external influences on citations

Evidence on the declining relevance of citations

I find evidence that such external influences do play a role in determining both the level of relevance of backward citations (i.e. patents cited by the applicants) and the potential omission of relevant citations. It has been well documented that patent litigation has been rising over time Marco et al. (2017). The number of backward citations (excluding self-cites) made by new patents has also increased, more than doubling from 2.3 to 6.0 over the period 1985-2015 (B.1).[5] Meanwhile, the average similarity of patents to their backward citations has declined (from 0.28 to 0.25, B.4) as well as the percentage of citations made to patents in the same primary class (54.1% to 34.4%, B.3). The decline in similarity to citations is robust across citations from (i) the same and different primary class; (ii) the same and different cities (see B.5,B.6). Taken together, these trends would indicate that the relevance of citations have been diluted by the addition of less related citations. However those made to patents within the same MSA has increased, although not consistently over the period: the share of local backward citations rose from 9.3% in 1985 to 12% in 2015.

Evidence of external influences on rate of local citations

Sample construction I examine the possibility of strategic omission of relevant citations using a dataset of "potentially citeable" patent pairs. I sample of set of *target* patents and find a complete list of their backward citations. For each backward citation, I find all their forward citations: each target patent is then matched with another such forward citation, granted *after* the target. Thus, each target is matched with a patent that has a backward citation in common, so that the target is "potentially citable" by the matched patent. I then calculate cross-patent similarity for each pair. To prevent

[5]To avoid truncation bias, only citations granted within the previous 10 years of the new patent were counted.

noise from bins with few observations[6], the lowest bin includes all values below, and the highest bin includes all values above. Over 2.4 million pairs of similarities are calculated.

Evidence of strategic omissions In the absence of strategic motives, the rate of citation should be increasing monotonically with similarity between patents. Greater similarity between the texts of two patents should indicate greater potential relevance. Overall, I find that the rate of citation is *not* increasing monotonically with similarity; the rate of citation in fact declines for patent pairs that have the highest level of mutual similarity. While 6.3% of target patents are directly cited when their similarity ranges between 0.5-0.6, only 4.2% are cited for similarity 0.6+. To account for technology differences, I find that this trend also holds for patent pairs within the same primary class: 7.8% of target patents are directly cited when the patent pairs have similarity between 0.5-0.6, and only 4.6% when similarity is 0.6+. (See B.7,B.9,B.1) In fact, the only sample group for which the rate of citation *does* increase monotonically is for patent pairs in the same city, which confirms the lack of incentive for strategic omission (B.8). This is contrasted by the stark decline in the rate of citation for patent pairs from different cities with the highest similarity: while 6.3% of target patents are cited when similarity is 0.5-0.6, only 2.2% are cited when similarity is 0.6+. For patent pairs in the same city, the rate of citation increases from 6.3% to 7.5%. Interestingly, the convergence of the rate in citation up to the 0.5-0.6 bracket might indicate diminishing strategic incentives to omit non-local patents as patents become more similar, but the divergence in their citation rates for patents with the highest similarity strongly indicates that firms are strategically leaving out the most relevant citations to patents from other cities. Local patents also over-represent less relevant citations, as the citation rate for pairs with lower similarity are consistently higher for local patent pairs. These findings taken together provide evidence that external influences on the selection of citations tends to favour local citations overall.

4. Inventor mobility and knowledge spillovers

Inventors are expected to be consistent in their knowledge of their own prior patents and prior citations. This fact can be exploited to further explore the nuances of the citation measure of knowledge flows. I examine the rate of citation for their own previous work, to see if citations may “miss” existing knowledge flows due to strategic motives after the inventor changes firms. Further, I compare the lists

[6] Below the 1st percentile and above the 99th

of citations made by inventors before and after they change firms to see how much of a difference this makes in their reported knowledge flows.

Finally, previous research such as Almeida and Kogut (1999); Azoulay et al. (2011) have used changes in the citation rate once inventors move cities to argue for localization. I compare the similarity of the mobile inventor's patents to their new citations to determine if it impacted their firm's new innovation ouputs.

4.1. Rate of self-citation before and after firm change

A clear example of where strategic non-citation might emerge is in the rate of citation for inventor's own patents, once they move to a different firm. Since inventors cannot reasonably claim to be ignorant of their own inventions, we can safely assume that any discrepancies in the rate of citation must be attributed to strategic withholding on the part of the inventor or new firm. I compare the rate of inventor self-citation when they are at their first firm, to the rate of self-citation of patents at their second firm to the patents at their first firm.

Sample construction

Suppose inventor i has patents A, B, C at firm 1, and D, E at firm 2. Then I will compare the self-citation rates in the set AB, AC, BC before their firm change, and AD, AE, BD, BE, CD, CE after the change. Since inventors often work in slightly different areas at their new firm, it is also crucial to condition on pairwise similarity in order to ascertain the appropriate benchmark citation rate. I use all 12,377 inventors who have changed firms and their complete patents at their first and second firm to construct my sample. They account for 8.7% of the 141,583 total inventors in the data. I calculate the complete set of pairwise similarities in the resulting sample, which after removing outliers, results in a sample size of almost 3.3 million pairs.

Evidence of strategic omission

If, conditional on similarity, the rate of citation is lower for inventors after they change firms compared to before, then this indicates that the inventor or the new firm is more or less knowingly concealing relevant citations in order to enlarge the scope of the new invention. I find evidence to support this claim in figure 4.1 and table C.1. On average, prior to the move, inventors cite their own inventions at the first firm in 12.5% of the observations, while after the move this drops in more than half to 5.8%. To allow for the possibility that inventors switch firms in order to work in different technology areas, I then condition the rate of self-citation on the similarity between the inventor's own patents. While this rate increases sharply with similarity between the two inventions, there is gap in the rate before and after changing firms that is consistent and statistically significant at almost every level of similarity. The difference grows with similarity up to the 0.5-0.6 bracket, after which the two measures converge for the highest levels of pairwise similarity. In the similarity range 0.5-0.6, inventors prior to their move across firms self-cited at a rate of 26.5%, while after the change it becomes 17.4%, a difference of almost 9%. Interestingly, the difference is not statistically significant at the highest level of similarity, largely due to the tapering off of *within* firm self-citation. One explanation is that there is no risk of patent infringement lawsuits from yourself, and so firms can expand the scope of their new patents by not listing their own highly similar previous inventions. Finally, if inventors cited themselves at their new firms at the same rate as before their move, then the projected number of total citations would be 11,875, compared to the actual number of 6,118. This implies that there are almost as many "missing" self-citations as actual self-citations.

figure 4.2 shows an example of an inventor who produces two very similar inventions: first, US Patent 7204412 "Family store value card program" was assigned to CompuCredit Intellectual Property holdings, applied for December 27, 2004. US Patent 7325725 "Store value card account transfer system", assigned to Purpose Intellectual Property in February 5, 2005, and yet did *not* cite 7204412.

4.2. Changes in citations made before and after firm change

An implication of this finding may be that firms (that is, the assignee of the patent who "owns" the intellectual property), not inventors, determine which patents are cited in the application. Using the same sample, I then examine how many citations are shared before and after the inventor changes firms using the number and percentage of common citations as described in 2.3. Citations made to other patents assigned to the same firm are excluded prior to the analysis.[7] I find that changing firms

[7] Because outliers have an outsized effect in determining the average number of common cited patents, I drop observations with number of common cited patents above the 99th percentile. This drops approximately 3,264 observations.

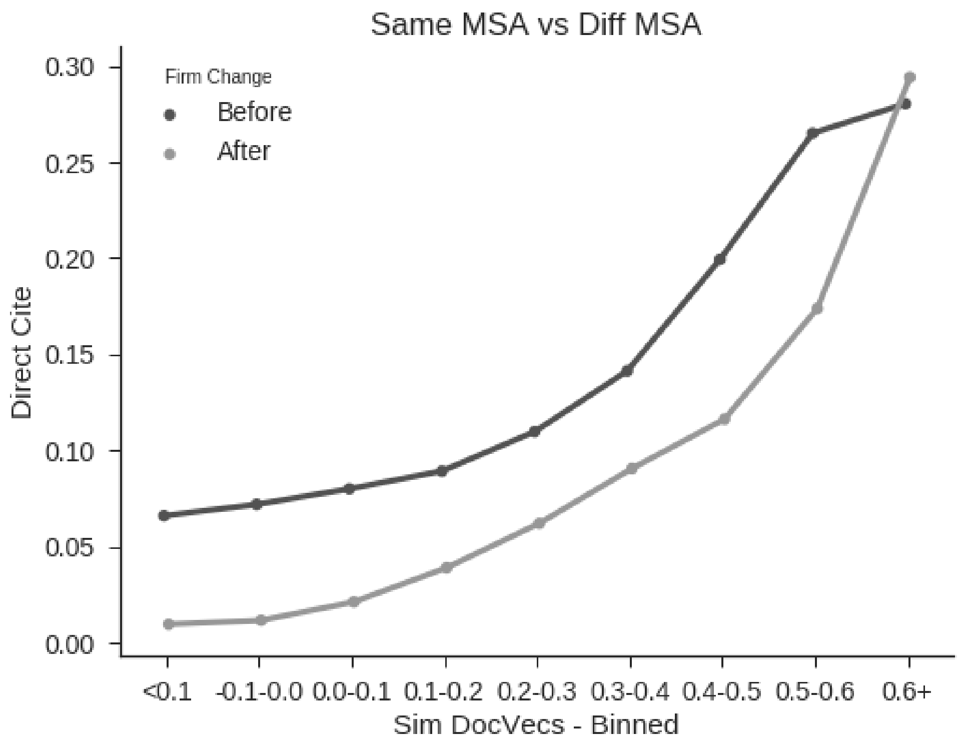


Figure 4.1: Rates of direct citation by DocVecs similarity. See C.1 for table of results.

significantly reduces the both the number and proportion of common citations made by the same inventor. Prior to the move, inventors on average shared 13% of backward citations with their own other patents. After the move, this drops to 5% overall: 6% if the inventor changed to a different firm in the same city, 3% if the inventor relocated to a different city. To account for the inventor changing innovation agendas once they switch firms, I also condition on similarity and find that a gap in the percentage of common citations exists for all similarity levels, and is particularly high when new patents have similarity of 0.5-0.6 to prior inventions. (C.1,C.2)

Looking just at the overlap backward citations, it would indicate that inventors are utilizing vastly different knowledge sources. But this is not accompanied by a drastic shift in their fields of interest. The change in similarity to their previous inventions is small, although significant: from a mean of 0.29 to 0.25 after changing firms. The number of pairs from the same primary class also reflect a smaller change in the inventor's output: from 35% prior to 30% after changing firms. There is a larger change for inventors who move cities as well, which indicates that inventors who switch firms and cities are altering their innovation agenda more drastically. While evidence suggests that inventors do change firms to produce different innovations, this change is slight compared to what is suggested

United States Patent	**7,204,412**
Foss, Jr.	**April 17, 2007**

Family stored value card program

Abstract

A family stored value card program is provided. One embodiment is a method for implementing a stored value card program. One such method comprises: identifying an existing stored value card account; and enabling a first customer associated with the existing stored value card account to establish a new stored value card account associated with a second customer, the new stored value card account linked to the first stored value card account.

Inventors: **Foss, Jr.; Sheldon H.** (Suwanee, GA)
Assignee: **CompuCredit Intellectual Property Holdings Corp. III** (Las Vegas, NV)
Family ID: **39360819**
Appl. No.: **11/022,739**
Filed: **December 27, 2004**

Prior Publication Data

United States Patent	**7,325,725**
Foss, Jr.	**February 5, 2008**

Stored value card account transfer system

Abstract

Systems, methods, computer programs, merchant terminals, etc. for transferring finds between stored value card accounts are provided. One embodiment comprises a method for loading a stored value card. One such method comprises: identifying a first stored value card account associated with a first customer; receiving a selection from the first customer of a second stored value card account associated with a second customer and a load amount for transferring to the second stored value card account; and initiating a funds transfer of the load amount from the first stored value card account to the second stored value card account.

Inventors: **Foss, Jr.; Sheldon H.** (Suwanee, GA)
Assignee: **Purpose Intellectual Property Management II, Inc.** (Las Vegas, NV)
Family ID: **39304185**
Appl. No.: **11/050,301**
Filed: **February 3, 2005**

Figure 4.2: Example of an inventor moving firms and not self-citing. US Patent 7204412 was not cited by US Patent 7325725.

	Num Common Cited	Pct Common Cited	Num Common Cited from Prev MSA	Pct Common Cited from Prev MSA	Sim DocVecs	Primclass Match
Before Firm Change, Mean	15.61	0.13	0.94	0.05	0.29	0.35
After Firm Change, Mean	1.92	0.05	0.05	0.02	0.25	0.30
p-value	0.00	0.00	0.00	0.00	0.00	0.00
After Firm Change - Same MSA, Mean	2.27	0.06	0.07	0.03	0.26	0.31
p-value	0.00	0.00	0.00	0.00	0.00	0.00
After Firm Change - Diff MSA, Mean	1.23	0.03	0.02	0.02	0.22	0.28
p-value	0.00	0.00	0.00	0.00	0.00	0.00

Table 4.1: Changes in number of common cited patents in inventor's own patents before and after firm change

by the change in their citations lists.

The discrepancy in the amount of overlap in the inventor's citation list and the similarity to their own previous inventions suggests that citations may be determined more by firm specific factors than the inventors themselves. An inventor may contribute a couple of citations they know and used before, and the rest is selected from a pool of citations that the firm uses, also likely influenced by their choice in lawyers. I find evidence in section 5.4 that lawyers play a large role in determining a patent's citations: patents from different firms that uses the same lawyer have 1.2-4.3% (or 0.8-2.9 S.D.s) higher proportion of common cited patents than pairs not sharing a lawyer (see results in table D.6). This is consistent with Wagner et al. (2014), who also show that firms who rely on professional service firms are more likely to cite patents that are part of the law firm's knowledge repository. These findings suggest that there is a further gap between what citations represent and the knowledge flows likely used by the inventor for their invention.

4.3. Effect of inventor mobility on patents in their new city

Following Almeida and Kogut (1999); Azoulay et al. (2011); Agrawal et al. (2006), I examine the changes in knowledge flows when inventors move cities. Of the 66,790 inventors I observe who changed firms in the previous subsections, about 12,846 inventors (19.2% of total) also moved cities. One key challenge with using mobility is that inventors often move cities to work in slightly different technology fields (as we saw above in 4.1). Thus, there may be an appearance in higher "knowledge flows" when in fact what is picked up is the inventor moving to a different city to work in a technology area that is concentrated in the new city. Adapted from Azoulay et al. (2011), who focus on academic citations made to mobile scientists, one way to partially control for this is to focus on knowledge flows from the inventor's patents *prior* to the move.

In Azoulay et al. (2011), they find that article-to-article citations from the scientists' new location increases markedly after the move. My findings corroborate this pattern. I focus on the 6,497 prior patents from inventors who moved cities. Each patent has received at least one (non-self) forward citation. On average, 2.91% of these citations matched the new location before the inventor moved. Afterwards, this rate jumps to 7.75% (p-value$= 0.00$). Once again, citations provides unequivocal evidence for localization.

However, does the citation represent a knowledge flow from the mobile inventor's patent, or merely that the firm now knows the mobile inventor? That is, is there evidence that firms in the mobile inventor's new city who cite their prior patents are actually influenced by these patents, or is the citation in some ways "perfunctory", reflecting the inventor's reputation or part in the local inventors' network rather than a knowledge spillover from the inventor to the firm.

Sample construction

For the prior patents of mobile inventors, I gather all citations that were made by firms in the new city that had *not* cited the inventor before. Prior to their move, 4,316 assignees from the new location had cited their past work. After the move, this number jumps to 10,578, with new citing firms accounting for 80.3% of the total. I focus on these firms as it is somewhat plausible that they have newly "discovered" the mobile inventor's work due to the inventor's presence in the city. These new citing firms make 27,817 forward citations to the mobile inventor's prior patents. For each forward citation, I try to select a control patent from the same primary class and firm, granted as close as possible in date, that does *not* cite the same prior patent. I only succeed in finding a control patent in 8,951 cases as not all firms have prior patents in the same primary class, or any prior patents at all.

Evidence of knowledge flow from citation to newly citing firm vs "perfunctory" citations

I attempt to gauge the relevance of the mobile inventor's prior patent on the newly citing firm in two ways. First, I compare if the citing patent is more similar to the prior patent compared to the control. Then, I compare the average similarity of the citing patent and the control to their firm's own prior patents in the previous 5 years (i) overall; (ii) within the same primary class. This is to determine whether or not the citing patent represents a departure from the firm's usual innovation agenda, due

	Sim DocVecs to Cited	Mean Sim Docvecs, Own Prior Pats	Mean Sim Docvecs, Own Prior Pats in Citing PC
Citing	0.278	0.281	0.328
Control	0.234	0.28	0.327
t-value	26.637	1.096	1.458
p-value	0.00	0.273	0.145
N	8951	6407	6338

Table 4.2: Changes in number of common cited patents in inventor's own patents before and after firm change. Differences in the number of observations arise due to a lack of other prior patents for citing patents' firms in different categories.

to the influence of the new inventor's knowledge flow to the firm.

I find that the citing patent is more similar on average to the cited prior patent compared to the control. In 4.2, the mean similarity of the citing patent is 0.278, while mean similarity is 0.234 for the control. The citing patent is about 18.8% more similar to the cited patent. However, I also find evidence that the citing patent is in any ways a "departure" from the firm's usual inventive activities. The citing patent and the control have identical average similarity to their own firm's prior patents, both across all primary classes and within the same primary class. When I rank the similarity of citing firm's prior parents to the new inventor's patent, I find that the citing patent was most similar in approximately 30% of cases. The median rank of the citing patent is 2.

These results suggest that while the new inventor may have influenced the citing patent, this patent was produced within the existing agenda of the citing firm. It is consistent with the explanation that firms in the new city were already working within the new inventor's technology field, and are citing the inventor who has become a peer. This suggests that those with the "absorptive capacity" (Cohen and Levinthal (2000)) to appropriate the knowledge brought by the new inventor are largely working within the same domain. As to whether or not the existing firm would have made the same invention *but for* the knowledge flows from the new inventor, the evidence is unclear. Some influence is suggested, but perhaps the contribution is not significant enough to drastically alter the invention.

5. Estimation of localization

Similarity and citations measure knowledge flows in different ways. I now turn to addressing the central question of whether there is geographic localization in knowledge spillovers, where I proxy for knowledge spillovers using both citations-based measures and cross-patent similarity. The general empirical specification has the form:

$$KS_i = \beta_0 + \beta_1 I(MSAMatch)_i + X_i + \epsilon_i \tag{5.1}$$

Where KS_i represents the measure of knowledge spillover for observation i, $I(MSAMatch) = 1$ if the knowledge spillover comes from the same city or MSA as i and 0 otherwise, and X_i represents controls for $i$including time effects and technology field effects, which will be discussed further in section TODO. In order for the various measures of knowledge spillovers to have some comparability, I normalize all measures to have a mean of 0 and standard deviation of 1.[8] The estimate for β_1 will then represent the effect of localization on knowledge spillovers, in this case how many standard deviations more knowledge spillovers local patents receive compared to non-local patents.

5.1. Benchmark: Extension of JTH (1993)

I replicate and extend the work of JTH in order to have a baseline to compare the magnitude of localization effects. JTH sampled patents in their control (target) group in the following manner: from the years 1975 to 1980, they select a random sample of Top Corporate (top 200 by R&D total expenditure measured by Compustat) and Other Corporate patents, and all patents granted to Universities. Their sample size is 950 for 1975 and 1450 for 1980 respectively. Then, for each "target" patent in the sample find a control patent that is as close as possible to the target in *grant date* in the *same patent primary class.* JTH claim that this accounts for the "existing distribution of technological activity," and thus if forward citations are more likely to be from the same geographical area as the target patent over the control, then it is evidence for the existence of localized knowledge spillovers.

In my method, I use a larger sample of target patents granted 1976-2005,[9] and limit forward citations to be within 10 years of the target patent's grant date. Self-citations of patents granted to the same assignee are similarly excluded. The only point of departure is that due to lack of data, I do not use separate categories of patents by assignee "type", and pool all patents by grant year. Compared to the original JTH results (table III, p. 590), my results are fairly well aligned with their 1980 cohort figures for top corporate patents. [10]

[8] Estimates for raw measures can be found in the appendix.

[9] 2005 is the last year that 10 year forward citations are available for

[10] 8.8 for target match and 3.6 for control match; compared to 9.09 and 3.77 for my results. Slight discrepancies may arise due to sample selection and slight differences in removing self-citations.

Summary table

I find that the measure of localization using citations based spillover measures is substantial and *rising* in size over time, concurring with Sonn and Storper (2008). The difference in the percentage of citations matching the target's MSA grows from 5.32% in 1975-85 to 6.46% two decades later. Relatively speaking, the effect is substantial: local patents are twice as likely to cite other local patents as non-local patents of the same primary class. The increase in both the target and the control's citations matching the target's MSA may indicate growing concentration in relevant inventions across locations.

Grant Year	Pct Targ Cites in MSA_T	Pct Control Cites in MSA_T	Diff	t-stat	p-value	N obs
1975-85	9.09	3.77	5.32	33.01	0.0	34489
1985-95	9.71	3.48	6.24	55.25	0.0	59102
1995-05	10.98	4.52	6.46	71.97	0.0	99248

Table 5.1: JTH Extension Results

Identifying Localization

To isolate the effect of localization, the above exercise can be represented as a regression model in the form equation (5.1):

$$pct\,cites\,in\,MSA_{T,i} = \beta_0 + \beta_1 I(MSA_i = MSA_T) + X_i + \epsilon \tag{5.2}$$

Where $i \in \{T = target, C = control\}$. Here, if patent i is the target patent, the indicator $I(MSA_T = MSA_T) = 1$, while for the control patent $I(MSA_C = MSA_T) = 0$. In this case, $\hat{\beta}_1$ represents the increase in the (standardized) percentage of local citations when the patent is also local, compared with non-local patents. The controls X_i include year fixed effects for when patent i was granted, and primary class fixed effects, which indicates the primary class attributed to patent i. While many other fixed effects could be included, the original study does not use any and these results are intended to be closely comparable with those estimates. Further controls will be included in 5.5.

The main concern with this specification, as discussed in Jaffe et al. (1993); Azoulay et al. (2011); Belenzon and Schankerman (2013) amongst others, is that the location of the patent is endogenous.

Patent i may be located in a particular city in order to take advantage of knowledge spillovers, and other agglomeration externalities, from the concentration of other firms also innovating in i's particular technology field. Since there is likely to be positive causal effects in both directions,[11] the sign of the simultaneity bias is likely to be positive (Wooldridge (2015)). As a first step towards remedying for technology field effects, primary class fixed effects are included for each patent. For pairwise spillover measures, the similarity in their respective technology fields is included as a better proxy for technological proximity, as discussed in 2.3. Additionally, evidence from 3 indicates that strategic considerations may lead to over-citation of local patents. Thompson (2006), on the other hand, finds that examiner added citations are less likely to be local than inventor's own, which may attenuate the over-estimation of localization. Results for 5.2 are presented in 5.3.

5.2. Changing the spillover measure to similarity

In devising an analogous similarity measure of spillovers using the same sample above, I consider what approach is possible under similarities that was not using citations, that might better approximate local spillovers. The use of the control patent has been scrutinized previously, notably by Thompson and Fox-Kean (2005), who argue that technological differences cannot be fully accounted for by selecting on primary class. With similarities, we do not necessarily require a control patent as a point of comparison. Instead, I take the cross patent similarity of a target patent with citations from its own MSA, to citations in different MSAs. In this approach, localization is present if a patent is more similar to its forward citations from the same MSA compared to those from different MSAs. The interpretation is that local patents influence local citations more; or analogously, that more relevant knowledge flows arise from other local patents.

Summary table

I find that target patents are more similar to citations matching its own MSA, compared to citations from other MSAs. The average similarity of citations in the same MSA is 0.35, compared to 0.29 for citations in other MSAs (see 5.2). However, the relative magnitude of the effect is much smaller compared to the JTH citations measure. DocVecs measures find that local citations are about 15% more similar; once the measure is normalized, the effect decreases substantially (see 5.3).

[11] As long as the size of the effect is less than 1, which is highly likely to be true: higher knowledge spillovers should not increase the likelihood of being in the same location by a factor greater than 1, and vice versa.

Grant Year	$\overline{sim_{DV}(T,j \mid j \, in \, MSA_T)}$	$\overline{sim_{DV}(T,j \mid j \, not \, in \, MSA_T)}$	Diff	t-stat	p-value
1975-85	0.35	0.29	0.05	25.53	0.0
1985-95	0.33	0.29	0.05	33.05	0.0
1995-05	0.32	0.28	0.04	42.39	0.0
2005-15	0.31	0.27	0.04	26.92	0.0

Table 5.2: JTH Replication Results with Similarity Measures.

Identifying Localization

In regression form, I estimate:

$$\overline{sim(T,j)} = \beta_0 + \beta_1 I(MSA_j = MSA_T) + X_T + \epsilon \tag{5.3}$$

Here, j is a patent that cites target T. $\overline{sim(T,j)}$ is the *average* of the pairwise similarities between the target and each of the forward citations j that are either located in the same MSA as T or not. $\hat{\beta}_1$ measures the increase in the (normalized) average similarity of a patent to citations in the same MSA compared to citations in different MSAs. Controls X_T are year and primary class fixed effects.

The issues with identification largely echoes what was discussed in the previous section. If firms with more similar technological innovations are more likely to collocate, then average citation similarity will appear localized. As discussed in section §3, exogenous influences may induce an over-representation of relevant local citations. Results for equation (5.3) are presented in table 5.3.

5.3. Measuring knowledge spillovers using similarity across random patent pairs

I abstract away from the baseline methodology in order to account for knowledge flows that may occur outside of direct citations. I change the measure of spillovers to similarity across random pairs of patents within a technology field, that is either (i) within primary class or (ii) within a NAICS field. [12]This approach is intended to capture broader relationships in knowledge shared within technological fields. The concept of knowledge spillovers has been critiqued by Breschi and Lissoni (2001) for

[12]The caveat is that the interpretation of the localization measure is different to those using primary class patents, as in this case it measures the effect of being in the same location for two patents within the same *industry*, which is a much wider definition of technology. The mean cross-patent similarity is much lower in this sample, which leads to larger estimates for coefficients on the log transform, even if the absolute magnitude of the change is roughly constant.

being not well defined, as they are supposed to reflect “information externalities” that arise from non-market interactions. Knowledge spillovers may be better defined by illustration: one way to portray the effect is in the fact that increase in firm R&D leads to a rise in sectoral productivity, perhaps through the firm’s contribution to the common knowledge of the sector. An interpretation of patent similarity is precisely that it measures the shared knowledge (from other patents and external sources) between two patents. Thus, higher similarity between local patents would indicate that a *greater degree of common knowledge* is shared between patents from the same city, after controlling for their technology fields, which fits well with the definition of localized knowledge spillovers.

The use of both within-NAICS and within-primary class samples allows me to compare spillover effects at broader and narrower technology levels. If localization measure is different for NAICS and primary class, then this indicates that there are differences in the dynamics of knowledge diffusion at the industry level and at more specialized or granular definitions of technology. In addition, using the NAICS sample which is comprised of numerous primary classes allows me capture the possibility of knowledge spillovers occuring *across* narrower subfields of technology.

I begin by measuring localized knowledge spillovers in a comparable way to the JTH-based methodology described above. Since the sample used in this exercise is different to the sample in the previous section, I add a citations based “similarity” measure in order to have a more commensurate spillover measure to compare similarity with. One approach of measuring similarity using citations as outlined in section 2.3 is to find the proportion of patent i’s backward citations that appear as backward citations for both i and j ($pcc(i, j)$ in equation (2.3)). This measure can be thought of as a citations-based proxy for the degree of similitude between the knowledge flows to patents $i$and j. In this case, localization is significant if the proportion of common cited patents is higher between patent pairs from the same location.

Sample construction I sample a set of target patents, and pair each of these patents with a random patent within the same field. Patents from the same MSA are slightly over sampled to ensure a sizable number of patent pairs from the same MSA across a range of technological fields. Patent pairs are granted within 5 years of each other and are assigned to different firms. While some patent pairs may have the same target patent, the number of appearances made by multiples of the same patent is extremely small relative to the entire sample, thus curtailing the presence autocorrelation.[13]

[13]Heteroskedastic-robust standard errors are used in regression estimates

Identifying Localization

In regression form, I estimate for each technology field sample:

$$sim(i,j) = \beta_0 + \beta_1 I(MSA_i = MSA_j) + X_i + \epsilon_{i,j} \tag{5.4}$$

$$pcc(i,j) = \beta_0 + \beta_1 I(MSA_i = MSA_j) + X_i + \epsilon_{i,j} \tag{5.5}$$

X_i represents grant year and primary class fixed effects for patent i.[14] There are simultaneity bias concerns here between $sim(i,j), pcc(i,j)$ and location match $I(MSA_i = MSA_j)$ as well, if more similar patents are also more likely to collocate. However, as discussed in 5.1, the sign of the simultaneity bias is highly likely to be positive, which means localization will be overestimated under this specification. Further controls for pairwise knowledge spillovers are included in (TODO section) below.

5.4. Regression Results

I find that the citations-based measure of the JTH extension exercise finds much larger relative effects for localization that for all other measures. Being in the same location increases the percentage of citations from that location by 0.24 standard deviations in 1975-85; this grows to 0.28 in 1995-2005. [15] By comparing same MSA citations to different MSA citations, a very different picture emerges. The normalized measure of average similarity does not find significantly higher similarity for in-MSA citations (although the difference is positive and significant in the raw data, table D.1) for the years 1985-2005.

For the within-NAICS sample, localization estimates using DocVecs similarity finds that being in the same and location increases similarity across patents by 0.03-0.06 standard deviations. This was lowest in 1975-85, rose 1985-2005, before declining slightly in 2005-15. For within-Primclass, the time trend is roughly similar: lowest at 0.05 in 1975-85, increasing to 0.08 1985-2005, and declining to 0.06 in 2005-15. Localization effects are found to be higher when spillovers are measured using proportion of common citations (pcc), although still much lower than in the benchmark comparison.

[14] Fixed effects for patent j are excluded due to dimensionality issues. Since i,j are more or less symmetric, this does not affect results.

[15] Results 2005-15 excluded as forward citations data only extend to 2015

KS = Pct Cites in Target's MSA, JTH Sample			
	1975-85	1985-95	1995-05
$I(MSAMatch)$	0.2416***	0.2846***	0.2966***
	(0.0073)	(0.0051)	(0.0041)
N	58647	107358	185154
Adjusted R^2	0.03	0.05	0.05
Year FE	True	True	True
PC FE	True	True	True

KS = Similarity to Citations, JTH Sample			
	1975-85	1985-95	1995-05
$I(MSAMatch)$	0.0444***	0.0158	0.0064
	(0.0145)	(0.0097)	(0.0066)
N	38541	69612	122217
Adjusted R^2	0.03	0.03	0.04
Year FE	True	True	True
PC FE	True	True	True

Table 5.3: Baseline regression results for JTH Sample.

For the within-NAICS sample, local patents have 0.03-0.09 standard deviations higher pcc. For patent pairs within the same primary class, local pairs have 0.12-0.18 higher pcc compared to non-local pairs. Though both measures are constructed using separate data, they are aligned in overall time trends: lowest in 1975-85, higher in 1985-2005, declining in 2005-15. Both measures also finds higher localization effects for patents within the same primary class, which indicates that local knowledge is more appropriable in narrower definitions of technology. The results are consistent with raw data, reported in table D.2. Even before controlling for technological proximity, there is already evidence that local knowledge spillovers are much smaller than prior studies suggest.

KS = Sim DocVecs, Within-NAICS				
	1975-85	1985-95	1995-05	2005-15
$I(MSAMatch)$	0.0323***	0.0605***	0.0610***	0.0571***
	(0.0053)	(0.0043)	(0.0034)	(0.0030)
N	192841	281222	437685	569252
Adjusted R^2	0.07	0.07	0.08	0.06
Year FE	True	True	True	True
PC FE	True	True	True	True

KS = Sim DocVecs, Within-Primclass				
	1975-85	1985-95	1995-05	2005-15
$I(MSAMatch)$	0.0527***	0.0798***	0.0787***	0.0573***
	(0.0064)	(0.0050)	(0.0038)	(0.0032)
N	170882	252174	401623	529686
Adjusted R^2	0.07	0.07	0.08	0.07
Year FE	True	True	True	True
PC FE	True	True	True	True

KS = Pct Common Cited, Within-NAICS				
	1975-85	1985-95	1995-05	2005-15
$I(MSAMatch)$	0.0327***	0.0777***	0.0875***	0.0725***
	(0.0056)	(0.0076)	(0.0068)	(0.0051)
N	188734	273714	384782	552981
Adjusted R^2	0.00	0.00	0.00	0.00
Year FE	True	True	True	True
PC FE	True	True	True	True

KS = Pct Common Cited, Within-Primclass				
	1975-85	1985-95	1995-05	2005-15
$I(MSAMatch)$	0.1188***	0.1755***	0.1524***	0.1379***
	(0.0097)	(0.0095)	(0.0070)	(0.0056)
N	167005	245355	352428	514322
Adjusted R^2	0.01	0.01	0.01	0.01
Year FE	True	True	True	True
PC FE	True	True	True	True

Table 5.4: Baseline regression results for Number of Common Cited and Sim DocVecs.

Validation exercise: regression results for matching on other variables

One concern may be that instead of geographic proximity being a weak determinant of patent knowledge relatedness, in fact similarity is not a good measure of patent relatedness due to noise. I address this concern by seeing if other match variables are strong determinants of similarity across patents. Using prior expectations, I expect the following match variables to have a significant effect on patent similarity: $I(Lawyer\,Match)$, if patents are assigned to firms that share the same lawyer; $I(Inventor\,Match)$, if patents share an inventor (after inventor relocates to different firm); $I(Primclass\,Match)$, if patents are from the same primary class; and $I(Common\,Cited \geq 1)$ to indicate the presence of at least one common cited patent between the pair. The results for DocVecs similarity are reported in table D.3, and for pcc in table D.4.

I find that estimates for these other match effects are large and significant. For DocVecs, the estimated effect of sharing an inventor increases similarity by 1.27-1.52 S.D.s; by 6.5-13.5 S.D.s for pcc. Sharing a common cited patent increases DocVecs similarity by 0.84-1.51 S.D.s; by 8.9-16.9 S.D.s for pcc (this is "mechanical" as the indicator is highly correlated with pcc). The rise in citation rates in recent decades has meant that sharing a common cited patent has declined in effect on DocVecs similarity, consistent with evidence in section §3. The effect of lawyers has a much smaller effect on similarity (increases by 0.28-0.45 S.D.s) compared to pcc (increases by 0.8-2.9 S.D.s). While this indicates that lawyers do have a significant effect in both patent text and patent citations, the citations measure is much more affected by the presence of a shared lawyer, indicating that firm fixed effects play a larger role in determining this measure of knowledge spillover. Ideally, we would not want external factors such as lawyer choice to affect our proxies for knowledge spillovers; explicitly controlling for them will be important in this case. For primary class match, the effect is reversed: larger for similarity (increase by 0.4-0.44 S.D.s) than for pcc (increase by 0.12-0.22). Overall, these findings indicate that the proportion of common cited patents is more sensitive to inventor and firm specific factors than technology, compared to similarity. This is due to the fact that pcc is 0 or close to 0 for most observations so that the size of its standard deviation is small, and are drastically higher for patents that share an inventor or lawyer. These estimates do show that similarity (as well as pcc) is fully capable of finding large and significant effects and that low localization estimates are not driven by the presence of noise.

5.5. Extending technological proximity controls

There has been much discussion in the existing literature as to whether or not selecting within a technology field is an adequate control for technological proximity. In JTH, the problem is summarized as: "... if a large fraction of citations to Stanford patents comes from the Silicon valley, we would like to attribute this to localization of spillovers. A slightly different interpretation is that a lot of Stanford patents relate to semiconductors, and a disproportionate fraction of the people interested in semiconductors happen to be in the Silicon valley, suggesting that we would observe localization of citations even if proximity offers no advantage in receiving spillovers." Knowledge overlap may occur just through being in the same or similar technological fields; that the knowledge inputs to creating innovation in class *396: Photography* might overlap significantly with class *398: Optical communications*. If these two fields are collocated (i.e. innovation in both classes occur in the same location), then we may over-attribute the effect of technological affinity to localization.

I also use cross-field similarity and cross-location-field similarity as specified in equation (2.4), equation (2.5) to improve the control for technological proximity. These measures are precisely informative about the past knowledge affinities between patents across different primary classes. I control for technological proximity using field similarity only within the NAICS sample, as subfield controls are not available for primary classes. Thompson and Fox-Kean (2005) have previously used primary subclasses, but I find these measures to be extremely noisy: there are over 150,000 subclasses in the USPC system for approximately 2.3 million US patents, which on average implies just 15 patents per subclass.

Since each NAICS field encompasses only a number of primary classes, the number of outcomes for similarity across primary class may be limited. To increase variability, I also use MSA-specific primary class similarity as described in equation (2.5). This treats each primary class at each MSA as a separate subfield. The trade off, however, is that this measure can be much noisier, as many MSAs may only have a handful of prior patents in some patent classes. [16] These measures should also be a closer representation of the underlying patents, as it defines a narrow band of technology. In addition, high primary class-MSA similarity is more correlated with collocation between patents: patents from highly similar subfields are more likely to be located in the same city due to agglomerative benefits (shared labour pool, production inputs). For these reasons, this measure will further

[16] I remove the MSA-fields with less than 10 patents

limits the remaining knowledge spillover to be explained by "pure" location effects. The expectation is that localization will be lowest when MSA-fields are controlled for, and estimates when this control is included should be viewed as a lower bound. The inclusion of these measures also reduces the total sample size in by 18.5% for the within-NAICS sample, and 17.2% in the within-primary class sample, since not all patents have prior primary class-MSA observations.[17]

Additionally, there are other potential prior connections between patents that may make the similarity between patent pairs higher for non-knowledge spillover related reasons. For example, if the patent pair share an inventor (after inventor has changed to a different firm), we would expect to see much higher similarity between patents without the presence of a "spillover." If the patents are from different companies that share a lawyer, we may expect the lawyer to word inventions in a similar "style". I include all controls that may induce higher similarity for reasons other than knowledge flows.

Identifying localization

The regression model including further controls will be:

$$sim(i,j)_t = \beta_0 + \beta_1 I(MSA_i = MSA_j) + sim(pc_i, pc_j)_{t-5,t} + X_{i,j} + \epsilon_{i,j} \tag{5.6}$$

$$pcc(i,j)_t = \beta_0 + \beta_1 I(MSA_i = MSA_j) + sim(pc_i, pc_j)_{t-5,t} + X_{i,j} + \epsilon_{i,j} \tag{5.7}$$

Where pc_i represents the primary class of pc_i. With primary class-MSA similarity:

$$sim(i,j)_t = \beta_0 + \beta_1 I(MSA_i = MSA_j) + sim(pc_{i,MSA_i}, pc_{j,MSA_j})_{t-5,t} + X_{i,j} + \epsilon_{i,j} \tag{5.8}$$

$$pcc(i,j)_t = \beta_0 + \beta_1 I(MSA_i = MSA_j) + sim(pc_{i,MSA_i}, pc_{j,MSA_j})_{t-5,t} + X_{i,j} + \epsilon_{i,j} \tag{5.9}$$

Where pc_{i,MSA_i} represents patents in pc_i and also in MSA_i. t represents the grant year of patent i; j is granted within t to $t+5$. Each respective technological proximity similarity measure uses only the patents that were granted in the five years prior to t. Additionally, I include interaction effects between $I(MSA_i = MSA_j)$ and technological similarity. I expand the controls $X_{i,j}$ beyond year and primary class fixed effects to include: $I(Lawyer\,Match)$, $I(Inventor\,Match)$, $I(Primclass\,Match)$

[17]This does alter some estimates in the previous section, but within a range of 0.01 in difference.

and $I(common\,est\,inv)$, if the patents' assignees' (at the location of the patent) shared an inventor in the past ten years. Since the within-primary class sample is already matched at the primary class level, only the second set of regressions using primary class-MSA are available.

5.5.1. Regression Results

For the within-NAICS sample, the estimate of localization diminishes further with the inclusion of technology proximity controls. Using primary class similarity, the estimate of localization ranges from insignificant to 0.04 S.D.s above the mean; the interaction effect is insignificant. As expected, using primary class-MSA similarity produces estimates of localization that are slightly lower for each year, ranging from insignificant to 0.035 S.D.s. The interaction effect is found to be slightly positive and significant in 1995-2015. For the within-primary class sample, localization estimates also range from insignificant to 0.04 S.D.s once primary class-MSA similarity is controlled for; the interaction effect insignificant in this case. As within the fixed effects only results, the time trends are slightly different in each sample, with the NAICS sample showing a decline in localization from 1985-95, while the primary class sample shows a decline only from 1995-2005. In all cases, localization is found to be insignificant in 1980-85.[18] This may be partly driven by the smaller sample size in this year group.

[18] Due to a lack of data technology proximity measures are only available from 1980, as it requires patent data from five years prior.

NAICS Sample: Primclass Similarity								
	(1)				(2)			
$KS = sim_{DV}(i,j)$	1980-85	1985-95	1995-05	2005-15	1980-85	1985-95	1995-05	2005-15
$I(MSAMatch)$	0.0170	0.0390***	0.0300***	0.0274***	0.0178	0.0415***	0.0277***	0.0222***
	(0.0120)	(0.0070)	(0.0048)	(0.0038)	(0.0120)	(0.0069)	(0.0048)	(0.0043)
$I_{MSA} * sim_{DV}(pc_i, pc_j)$					-0.0012	-0.0045	0.0050	0.0085*
					(0.0116)	(0.0067)	(0.0047)	(0.0045)
$sim_{DV}(pc_i, pc_j)$	0.2805***	0.2913***	0.2816***	0.2932***	0.2808***	0.2925***	0.2804***	0.2908***
	(0.0090)	(0.0055)	(0.0040)	(0.0036)	(0.0095)	(0.0057)	(0.0042)	(0.0039)
N	40323	110982	215861	344313	40323	110982	215861	344313
Adjusted R^2	0.12	0.13	0.13	0.08	0.12	0.13	0.13	0.08
Year FE	True	True	True	True	True	True	True	True
PC FE	True	True	True	True	True	True	True	True
Inv & Lawyer Match	True	True	True	True	True	True	True	True

NAICS Sample: Primclass-MSA Similarity								
	(3)				(4)			
$KS = sim_{DV}(i,j)$	1980-85	1985-95	1995-05	2005-15	1980-85	1985-95	1995-05	2005-15
$I(MSAMatch)$	0.0109	0.0353***	0.0187***	0.0170***	0.0114	0.0357***	0.0202***	0.0194***
	(0.0121)	(0.0071)	(0.0048)	(0.0038)	(0.0124)	(0.0073)	(0.0051)	(0.0039)
$I_{MSA} * sim_{DV}(pc_{i,MSA_i}, pc_{j,MSA_j})$					0.0064	0.0040	0.0103*	0.0261***
					(0.0136)	(0.0079)	(0.0061)	(0.0057)
$sim_{DV}(pc_{i,MSA_i}, pc_{j,MSA_j})$	0.1259***	0.1259***	0.1643***	0.1830***	0.1245***	0.1250***	0.1618***	0.1764***
	(0.0070)	(0.0041)	(0.0035)	(0.0031)	(0.0076)	(0.0045)	(0.0038)	(0.0034)
N	40323	110982	215861	344313	40323	110982	215861	344313
Adjusted R^2	0.11	0.11	0.12	0.08	0.11	0.11	0.12	0.08
Year FE	True	True	True	True	True	True	True	True
PC FE	True	True	True	True	True	True	True	True
Inv & Lawyer Match	True	True	True	True	True	True	True	True

Table 5.5: Regression results including technology proximity and other controls with normed data for DocVec similarity.

Primclass Sample: Primclass-MSA Similarity								
	(5)				(6)			
$KS = sim_{DV}(i,j)$	1980-85	1985-95	1995-05	2005-15	1975-85	1985-95	1995-05	2005-15
$I(MSAMatch)$	0.0100	0.0248***	0.0316***	0.0157***	-0.0047	0.0353***	0.0413***	0.0216***
	(0.0143)	(0.0081)	(0.0054)	(0.0041)	(0.0213)	(0.0112)	(0.0075)	(0.0058)
$I_{MSA} * sim_{DV}(pc_{i,MSA_i}, pc_{j,MSA_j})$					0.0145	-0.0116	-0.0123	-0.0087
					(0.0175)	(0.0099)	(0.0077)	(0.0068)
$sim_{DV}(pc_{i,MSA_i}, pc_{j,MSA_j})$	0.0654***	0.0652***	0.0826***	0.0935***	0.0632***	0.0671***	0.0849***	0.0953***
	(0.0075)	(0.0045)	(0.0038)	(0.0035)	(0.0079)	(0.0048)	(0.0041)	(0.0038)
N	38324	106152	205248	324142	38324	106152	205248	324142
Adjusted R^2	0.08	0.08	0.09	0.06	0.08	0.08	0.09	0.06
Year FE	True	True	True	True	True	True	True	True
PC FE	True	True	True	True	True	True	True	True
Inv & Lawyer Match	True	True	True	True	True	True	True	True

Table 5.6: Regression results including technology proximity and other controls with normed data for DocVec similarity.

5.5.2. Comparing conditional means of similarity under varying levels technology proximity

To investigate the effect of technology proximity further, I then calculate the conditional sample mean of DocVec similarity for each bin of $sim_{DV}(pc_i, pc_j)$ and $sim_{DV}(pc_{i,MSA_i}, pc_{j,MSA_j})$ for the entire sample of observations. For the within-NAICS sample, the relationship between patent pair similarity and primary class similarity (D.1) appears to be linear and monotonic; however, some non-monotonicity emerges when conditioning the pair similarity on primary class-MSA similarity (D.2). Echoing results in 3, the mean similarity for pairs in the highest bin of primary class-MSA bin actually declines by 0.22 S.D.s, for both local and non-local pairs (p-value $= 0.00$ in both cases). Since high primary class-MSA similarity patent pairs are usually operating within the same or similar technology fields, this points to the novel finding that innovative firms are differentiating from rivals.

For within-primary class samples, a more striking pattern emerges: the mean similarity of local patent pairs is actually *lower* than that of non-local pairs, conditional on primary class-MSA similarity above -0.5 S.D.s. In table 5.7, the advantage to patent pairs rises as technological proximity within the two locations rises: for highly proximate primary class-MSAs, local patent pairs are on average 0.17 S.D.s lower in similarity compared to non-local pairs. A possible explanation for this phenomena may be that local firms are pursuing differentiated R&D agendas. Knowing more about other local firms' innovation might actually lead other local inventors to divert their research away from their rivals in order to both avoid potential infringement and widen the scope of their own patent. The fact that such an effect is not present for patent pairs within the same NAICS industry suggests that positive spillovers outweighs differentiation effects within broader technological fields; however this is reversed when looking within narrower bands of technology.

Regression results with lower similarity for local pairs in the same primary class I capture this relationship by estimating a segmented regression model for two lines: one where the norm $sim_{DV}(pc_{i,MSA_i}, pc_{j,MSA_j})$ is below 0 (i..e the mean); and one where it is above. I interact both terms with the MSA Match indicator. The specification is:

$$
\begin{aligned}
sim_{DV}(i,j)_t = \beta_0 + \beta_1 I(MSA_i = MSA_j) + \beta_2 I(sim_{DV,pc,MSA} \le 0) * sim_{DV}(pc_{i,MSA_i}, pc_{j,MSA_j})_{t-5,t} \\
+ \beta_3 I_{MSA} * I(sim_{DV,pc,MSA} \le 0) * sim_{DV}(pc_{i,MSA_i}, pc_{j,MSA_j})_{t-5,t} \\
+ \beta_4 I(sim_{DV,pc,MSA} > 0) * sim_{DV}(pc_{i,MSA_i}, pc_{j,MSA_j})_{t-5,t} \\
+ \beta_5 I_{MSA} * I(sim_{DV,pc,MSA} > 0) * sim_{DV}(pc_{i,MSA_i}, pc_{j,MSA_j})_{t-5,t} + X_{i,j} + \epsilon_{i,j}
\end{aligned}
\tag{5.10}
$$

Where controls remain the same as before. Regression results are presented in 5.8. I find that the slope of the regression line for local pairs is indeed lower than for non-local pairs when primary class-MSA similarity is above the mean ($\hat{\beta}_5 < 0$), but only for the decades 1995-2015. There is no significant difference in the two lines when primary class-MSA similarity is below the mean ($\hat{\beta}_3 = 0$). This indicates that local differentiation has played a role in innovation for the last two decades.

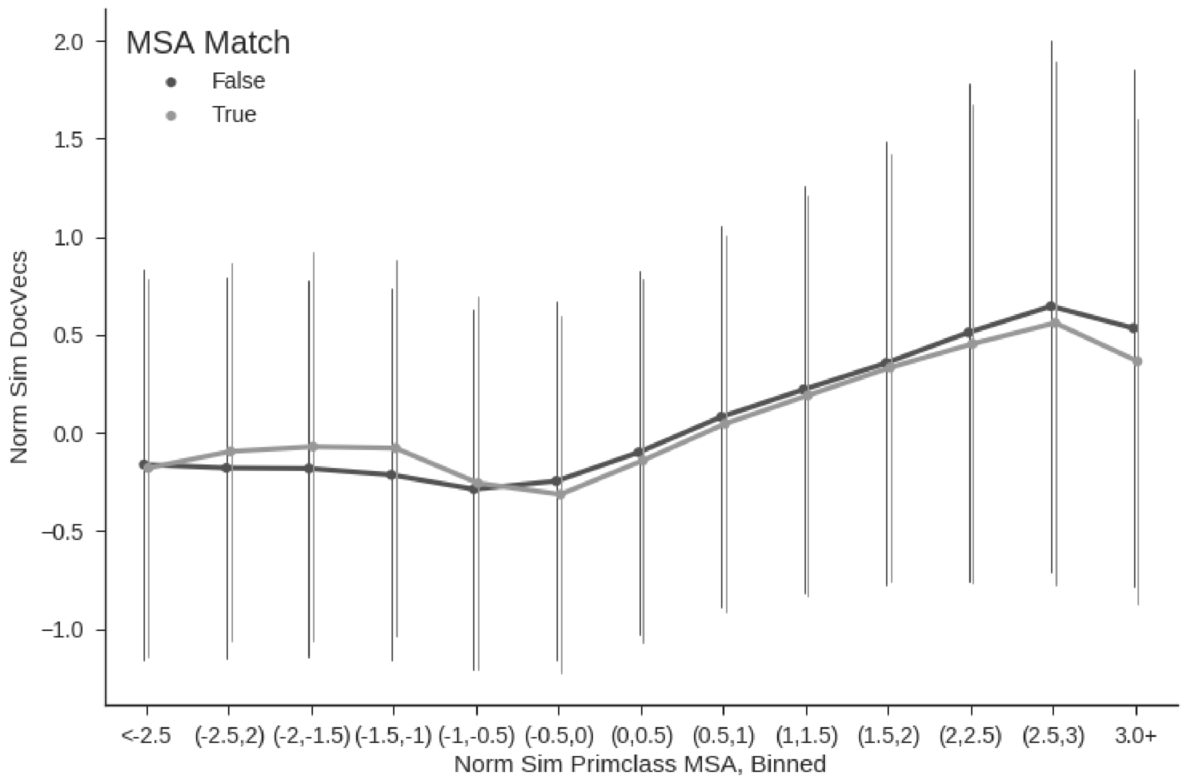


Figure 5.1: Conditional means of DocVec similarity by MSA-primary class similarity, within-primary class sample.

	<-2.5	(-2.5,-2)	(-2,-1.5)	(-1.5,-1)	(-1,-0.5)	(-0.5,0)	(0,0.5)	(0.5,1)	(1,1.5)	(1.5,2)	(2,2.5)	(2.5,3)	3<
Norm Sim DocVecs, MSA Match = T, Mean	-0.18	-0.095	-0.071	-0.078	-0.256	-0.314	-0.141	0.046	0.19	0.333	0.454	0.561	0.365
Norm Sim DocVecs, MSA Match = F, Mean	-0.162	-0.179	-0.182	-0.214	-0.287	-0.246	-0.1	0.081	0.221	0.356	0.512	0.646	0.534
MSA Match = T, N	129	115	262	455	1554	16874	73825	79380	47248	20507	9586	5750	7311
MSA Match = F, N	2348	2476	5443	13503	45109	151247	278745	196470	92429	37501	18525	9837	7674
Diff in Mean	-0.018	0.084	0.111	0.136	0.031	-0.068	-0.042	-0.035	-0.03	-0.023	-0.058	-0.085	-0.169
t-value	-0.2	0.913	1.769	2.978	1.257	-9.127	-10.85	-8.546	-5.245	-2.374	-3.727	-3.817	-8.119
p-value	0.841	0.363	0.078	0.003	0.209	0	0	0	0	0.018	0	0	0

Table 5.7: DocVec similarity conditional on primary class-MSA similarity for within-primary class patent pairs.

	1975-85	1985-95	1995-05	2005-15
$I(MSAMatch)$	-0.0085	0.0390***	0.0524***	0.0295***
	(0.0235)	(0.0128)	(0.0085)	(0.0063)
$I_{MSA} * I(sim_{DV,pc,MSA} > 0) * sim_{DV}(pc_{i,MSA_i}, pc_{j,MSA_j})$	0.0142	-0.0167	-0.0267***	-0.0212***
	(0.0196)	(0.0114)	(0.0089)	(0.0075)
$I_{MSA} * I(sim_{DV,pc,MSA} \leq 0) * sim_{DV}(pc_{i,MSA_i}, pc_{j,MSA_j})$	-0.1366*	-0.0299	0.0202	0.0118
	(0.0791)	(0.0470)	(0.0422)	(0.0406)
$I(sim_{DV,pc,MSA} > 0) * sim_{DV}(pc_{i,MSA_i}, pc_{j,MSA_j})$	0.0764***	0.0769***	0.1042***	0.1148***
	(0.0109)	(0.0067)	(0.0057)	(0.0051)
$I(sim_{DV,pc,MSA} \leq 0) * sim_{DV}(pc_{i,MSA_i}, pc_{j,MSA_j})$	0.0301*	0.0446***	0.0402***	0.0456***
	(0.0183)	(0.0105)	(0.0088)	(0.0084)
N	38324	106152	205248	324142
Adjusted R^2	0.08	0.08	0.09	0.06
Year FE	True	True	True	True
PC FE	True	True	True	True
Inv & Lawyer Match	True	True	True	True

Table 5.8: Regression results with primary class-MSA similarity with segmented regression for within-primary class patent pairs.

Example: Pharmaceutical drug patents from primary class 514 For example, consider the primary class-MSA *514-Philadelphia*. Class 514 refers to patents for pharmaceutical drugs. The primary class-MSA similarity for itself $sim_{DV}(514_{Phil}, 514_{Phil}) = 2.89$, while its similarity to *514-New York* $sim_{DV}(514_{Phil}, 514_{NY}) = 2.86$. However, the average similarity of patent pairs from *514-Philadelphia* to *514-Philadelphia* 1.13 S.D.s above the mean, while for pairs from *514-Philadelphia* to *514-New York* the average similarity is 1.34 S.D.s above the mean. This means that drug patents in Philadelphia are more similar to drug patents in New York than other patents in its own city, even if conditional primary class-MSA similarity predicts they should be roughly the same. What this may indicate is that the innovative activity of the pharmaceuticals industry is particularly sensitive to local rivals. Indeed, patent litigation is particularly active in this industry because "the pharmaceutical industry is one of the few industries that requires patent protection to ensure the profitability of its innovative products." (Avery (2008)) In fact, drug patents account for 29.4% of observations where primary class-MSA similarity is 2.5 S.D. or more above the mean. When I drop class 514 patents from the analysis (D.9), the significance of the difference disappears for primary class-MSA similarity in the range 2-3 S.D.s. This is illustrative of how an industry that strongly relies on patent protection might sustantially reduce positive externalities generated by local knowledge spillovers.

This finding is contrary to what is found using citations: in section §3, I discuss the possibility that increased patent litigation might also increase localization *in citations*, if the assumption that the probability of infringement discovery is higher amongst firms in the same location holds. Under JTH's methodology, *514-Philadelphia* patents cite other Philadelphia patents on average 29% of the time; while *514-New York* patents cite Philadelphia patents only 5% of the time. In this case, JTH would actually find a large and significant localization effects for pharmaceutical patents in Philadelphia. The two results are not necessarily logically inconsistent: both effects drive down the probability of patent litigation. Citations may be used for "litigation blocking" while differentiating patent text also reduces the potential for infringement. This suggests that patent texts may also exhibit some strategic qualities and it may be difficult to separate the effects of a modified research agenda from strategically differentiated patent text.

6. External sources of knowledge: new technology terms

The introduction of new terms into the patent corpus provides an opportunity to examine the influence of unobserved external knowledge on patent text. As new technology are developed, references make their way into patents. Patents that are the first to contain the new term are assumed to share some external sources of knowledge about the new technology. In order to verify if citations or similarity are able to reveal evidence of shared knowledge for newly appearing technological terms, I create a sample of patent pairs that were applied for in the first year that a new term appears. As before, patent pairs from the same firm are removed as within-firm shared knowledge do not constitute spillovers.

Hypothetically, firms appropriating the same new technology should share some knowledge sources in common. However, looking at their list of shared cited patents, even very similar new patents using the new technology share very few backward citations. Above the 99th percentile of DocVecs similarity for this sample (around 0.40), the number of shared cited patents is 0.087, which represents about 0.8% of the first patent's total cited patents. While the shared knowledge measure using citations is rising with similarity, the overall magnitude is still modest. In figure E.2, the percentage of shared backward citations rises more if the patent pair is local, which is consistent with the expectation that citation networks are more localized. This comes into effect after the 95th percentile similarity of around 0.3.

In table E.1, DocVecs similarity shows some indication of shared knowledge for patents using new technology. Citations based measures do not convey much relatedness between these patents. The distribution plots in figure E.3 for each measure of pairwise shared knowledge show that while DocVecs similarity is able to capture significant variation in the relatedness of new patents, citations measures do not.

Example: Adenovirus "Adenovirus" is a term for a virus that causes many common infections, particularly respiratory illness. With the development of gene therapy technology in the early 1990s, the first patent applications containing the term adenovirus appeared in 1993. Gene therapy delivers "correct" genes inside affected cells, and adenoviruses are often used as carriers for the corrected genes. In 1993, thirteen adenovirus patents were applied for that were later granted.[19] While all adenovirus patents apparently utilised some common external knowledge sources, the average number of backward citations that was shared was 0.03, which represented an average of 0.0% of backward citations made. figure E.4 displays a wide variety of similarity across the initial adenovirus patents, which had an average similarity of 0.26.

7. Policy Implications

My findings may also have important implications for R&D policy. Broadly speaking, there are three avenues for public support of private research: 1. funding for regional clusters; 2. funding for nationwide industry; 3. funding specific firms or institutions. In recent years, funding for regional clusters received a signficant boon in support from the Obama administration, who spent $225 million on

[19]Failed applications are not accessible via the USPTO.

regional cluster projects as of April 2012 Chatterji et al. (2014). The largest of these was the Energy Regional Innovation Cluster in Philadelphia, which received $122 million in funds from the Department of Energy. Local efforts include notably the Boston Waterfront Innovation District, launched in 2010 by Mayor Thomas Menino (Mehta et al. (2012)). Nation-wide industry R&D funding include support for renewable energy and nanotechnology (the National Nanotechnology Initiative was launched in 2001). While both efforts are underpinned by some economic rationale of generating positive externalities, the *kind* of externality differs: positive geographic externalities leading to better local economic outcomes for regional clusters, and positive social externalities leading to better population health for industry R&D funding.

The implications from my findings may be that, from the perspective of generating new innovation, the two forms of R&D funding are roughly equivalent. If there are less positive externalities generated by geographic proximity, then research boosted by local funds could be more easily appropriated by firms in the same technology field in other locations. On the other hand, this may weaken the argument for regional policy if it is intended to boost its own relative advantage in innovation. How much this evidence hampers enthusiasm for regional cluster policy depends on whether or not policymakers consider innovation-embodied knowledge spillovers to be a first order concern. Indications are that it is not; policymakers are usually much more focused on outcomes such as employment growth and rates of entrepreneurship. Furthermore, while the geographic localization of *innovative* knowledge may be challenged, there may be other forms of knowledge spillovers than are locally beneficial, such as knowledge about venture capital and supply chains, which can help local firms in other capacities.

The lower rate of geographically localized knowledge spillovers may be one possible contributing factor to the difficulty in "governmentally inducing" innovation. While notable successes such as SIlicon Valley and Route 128 in Boston receive considerable attention, many failed cases do not (Lerner and Seru (2015)). My findings show that firms are less inclined to share commercially appropriable innovative knowledge with local competitors. This may particularly affect the innovative capacity of firms in artificial "hubs" which lack the benefits from other forms of shared knowledge. If local knowledge spillovers are not particularly conducive towards innovation, then other agglomeration externalities such as shared labour pools and shared inputs may have particular importance in ensuring the survival of regional hubs. Unfortunately, these other factors may be much more difficult to engineer through policy interventions. Aritificial innovation hubs which rely heavily on knowledge spillovers generated through the presence of proximate firms may face challenges to success without the presence of other agglomerative benefits.

Finally, a separate but related literature examines the contribution of public funding for university research on innovation. I conjecture that if universities lack incentives to commercially appropriate their own research, it may indeed provide positive externalities to local firms engaged in similar innovation areas. Thus, one way that regional cluster policy could boost their relative advantage could be to also allocate funds for local university research.

8. Conclusion

This paper focuses on knowledge spillover dynamics of ideas embodied by innovation. By focusing on patent texts, the evidence for geographic localization is much weaker compared to results found by exclusively using patent citations. These findings reveal a potential difference between the the geographic dynamics of the patent-based inputs to the innovative process and outside influences, and further corroborates Jaffe et al. (2000) that some citations may not correspond to a knowledge link. This has some key implications for the existing literature that uses citations. First, citations can still be taken to represent inventor's and firm's knowledge of other inventor's and firm's patents. Thus, the validity of using citations as representative inventor networks still hold. Since the literature linking patent citation to patent value in some sense is derived from its importance to other inventors, this measure is not challenged by my findings. Second, I have so far only focused on similarity and citations between patent texts and not on academic publications. Roach and Cohen (2013) finds that citations made to non-patent references are less affected by strategic considerations on the part of the firm. Therefore, I would expect that citations from patents to publications and between publications act as better proxies for knowledge flows and the disparities between the citation and similarity measures should be smaller. This would be a fruitful avenue for future research.

Appendix

A. Text to Data

A.1. Text cleaning

Each abstract is stemmed to the root word (for example, computer to comput), and stop words (such as "and", "the") are removed. The first step in converting text to data is to represent words and documents in their simplest vector forms. For all algorithms besides Document Vectors, input into the algorithms involve the construction of a document-term matrix from all patents; each row is indexed by the document ID and each column represents a word in the vocabulary. A document row vector represents the count of the number of times the term appears in the document. For the terms, I drop all terms that appear in more than 10% of all patents, and those that appear in fewer than 20.[20] Of the resulting terms, I keep the most common 40,000, in order to maintain a manageable matrix dimensionality. Once all 2,306,041 patents have been transformed into a document-term matrix of dimension 2306041×40000, I proceed to transforming patents into a smaller dimensional vector representation using the methods described below. This procedure is commonly called the *bag-of-words* representation of text data.

[20] Including very common and very infrequent terms may introduce noise and considerable increases in computation times.

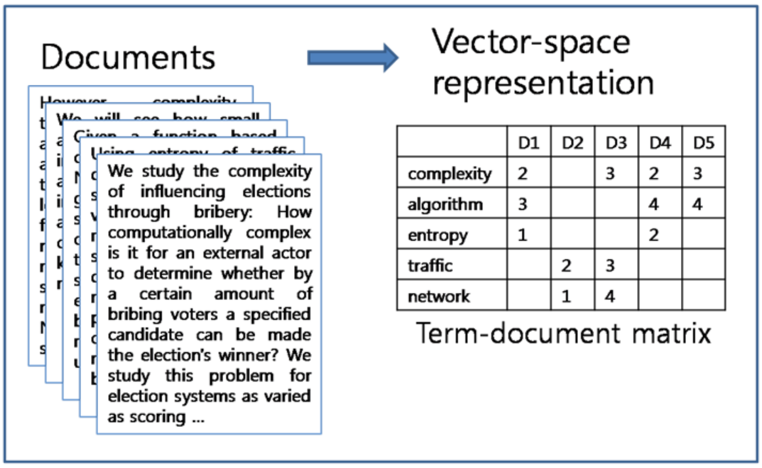


Figure A.1: Example of Document Term Matrix

A.1.1. Paragraph Vectors (Doc2Vec)

One recent advance in NLP which utilises neural networks is Paragraph Vectors, introduced by Le and Mikolov (2014). This is a straightforward extension of the word2vec model of Mikolov et al. (2013b,a). The word2vec model attempts to rectify one of the well-known problems of NLP: the inability of "one-hot" word vectors to account for word similarity. Typically, word vectors are represented as sparse vectors. For example, in a complete vocabulary of ["good", "fair", "fine"], the word *good* would be represented as the vector $[1,0,0]$, *fair* as $[0,1,0]$ and *fine* as $[0,0,1]$. Clearly, each of these vectors are orthogonal to each other and have a similarity of 0. Instead of using this class of word vectors, word2vec tries to represent words as dense vectors that encode such similarities; a word2vec vector for each of the three words ["good", "fair", "fine"] will have a *high* similarity.

The way that this is done is through looking at the *context* of a word. For example, for the sentence "Provides for unattended file transfers", the word "unattended" has the context ["Provides", "for", "file", "transfers"]. We want to represent each of these words as a vector of arbitrary dimension n. One way to account for context is to predict the context words given the target (Skip-gram); while another way is to predict the target word given the context (Continuous Bag-of-Words). Under Skip-gram, the optimization problem is to maximise the probability of any context word given the current center word. So the objective function is given by:

$$J(\theta) = -\frac{1}{V}\sum_{t=1}^{V}\sum_{-m\leq j\leq m}\log p(w_{t+j}|w_t) \tag{A.1}$$

Where θ represents all parameters: input vector ("one-hot") representation of each word, and the output word2vec representation of each word. m represents the length of the context window; for example $m=1$ gives the context for "unattended" as ["for", "file"]. The objective function is minimized using stochastic gradient descent.

Paragraph Vectors, or Doc2Vec, extends word2vec merely by adding an additional variable, which will be treated as an additional context vector: paragraph ID. For my data, this will be the patent number, which uniquely identifies every abstract document. Thus, including paragraph ID as an additional word for each context generated from that paragraph will also generate a unique vector associated with the paragraph, as well as the word vectors. Intuitively, the paragraph vector will represent what was learned in other context windows belonging to the paragraph, outside of the present context window: that is, it "acts as a memory that remembers what is missing from the

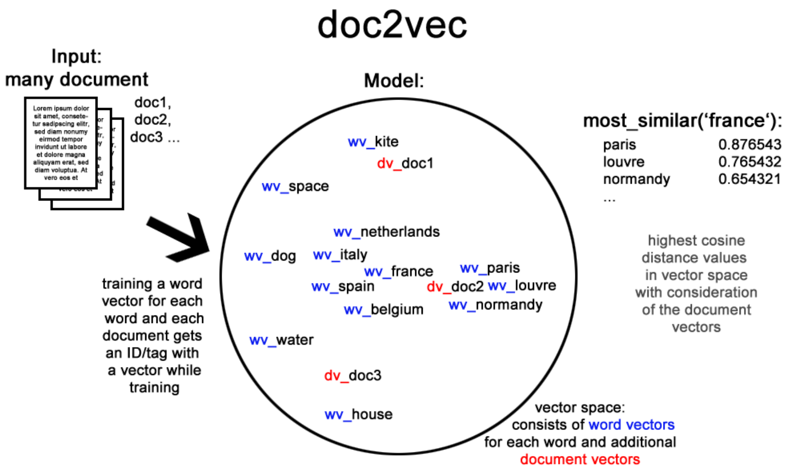


Figure A.2: Illustration of Document Vectors.

current context." (Le and Mikolov (2014))

Such an approach has been shown to be extremely powerful in accurately capturing cross-word and cross-document similarity (papers?), which is why it is the main focus of my analysis. Other vector representations of patents that I use do not specifically optimize to capture such similarity using contexts.

A.1.2. Latent Dirichlet Allocation

Latent Dirichlet Allocation, first introduced by Blei et al. (2003), is a method of Topic Modelling that assumes that a document can be represented as a linear distribution hidden variables called *topics*. It is a Hierarchical Bayesian hidden variables model. The Data Generating Process assumes that each topic is a linear distribution over terms in the corpus. For each document, which is a distribution over topics, each term is assumed to be generated by first drawing a topic, then drawing a term from that topic. Because this is an unsueprvised method, the algorithm then jointly determines the topics distribution over terms and each document's distribution over topics. See A.1.2 for more details on the assumptions of the LDA model. table A.1 shows a breakdown of selected topics' distribution over terms. figure A.3 provides an example of the input and outputs of the algorithm

The number of topics K is a parameter that is determined ex-ante; as per Hoffman et al. (2010), the recommendation is that the model with the lowest log perplexity be selected, although there is not a universally agreed upon procedure. I fit a LDA model on a training subset of the same document-term matrix representing all patent abstracts with $20, 30, ..., 120$ topics. Then, the model was fit on the test

Topic	Distribution over terms	Description
0	0.040*"network" + 0.039*"inform" + 0.033*"comput" + 0.031*"communic" + 0.028*"user" + 0.027*"memori"	Networks & Coding
2	0.066*"time" + 0.057*"sensor" + 0.040*"detect" + 0.032*"event" + 0.031*"paramet" + 0.027*"level"	Monitoring & Coding
11	0.116*"power" + 0.068*"voltag" + 0.049*"output" + 0.045*"circuit" + 0.026*"suppli" + 0.026*"transistor"	Electronics
36	'0.071*"composit" + 0.059*"polym" + 0.049*"weight" + 0.041*"coat" + 0.018*"resin" + 0.016*"c"	Polymers, Chemicals
53	'0.065*"metal" + 0.065*"solut" + 0.037*"ion" + 0.036*"carbon" + 0.032*"concentr" + 0.023*"reaction"	Metals, Chemicals

Table A.1: Selected Topics as outputted by LDA. Description added post hoc.

set and the log-perplexity calculated. I selected $K = 60$ as it had the lowest log perplexity across the models.

A snippet from the resulting topics is shown in A.1, alongside the six highest probability terms in teach topic. The output I am interested in is the probability across each of the 60 topics of each patent document. I take this as the Topic Model vector representation of each patent.

Data generating process With probabilistic models, treat observations as outcomes of a data generating model and infer the hidden parameters of that model using posterior inference. Define a "topic" as a discrete distribution over a fixed vocabulary. Assume each topic is generated by drawing a distribution over terms in the vocabulary represented by the vector: $\boldsymbol{\beta_k} = (\beta_{k,1},...,\beta_{k,V}) \sim Dir(\eta)$. Additionally, asssume that each document d is generated by the following process:

1. Draw a vector distribution over topics: $\boldsymbol{\theta}_d = (\theta_{d,1},...,\theta_{d,K}) \sim Dir(\boldsymbol{\alpha})$

2. For each word $w_{d,n}$:

   a) Draw a topic $k_{d,n} \sim Multinomial(\boldsymbol{\theta}_d)$

   b) Draw a word based on that topic's distribution over the vocabulary $w_{d,n} \sim Multinomial(\boldsymbol{\beta}_{z_d,n})$

Then the posterior of the hidden variables, conditional on the observed words in each document, is given by:

$$p(\beta_{1:K},\theta_{1:D},z_{1:D}|w_{1:D}) = \frac{p(\beta_{1:K},\theta_{1:D},z_{1:D},w_{1:D})}{p(w_{1:D})} \tag{A.2}$$

An inference algorithm is used to approximate the posterior. Thus, from the observed set of V vocabulary terms $w \in 1,...,V$, the hidden topics $k \in 1,...,K$ (a distribution over words in the vocabulary), and each document's distribution over topics $(\theta_{d,1},...,\theta_{d,K})$ are derived.

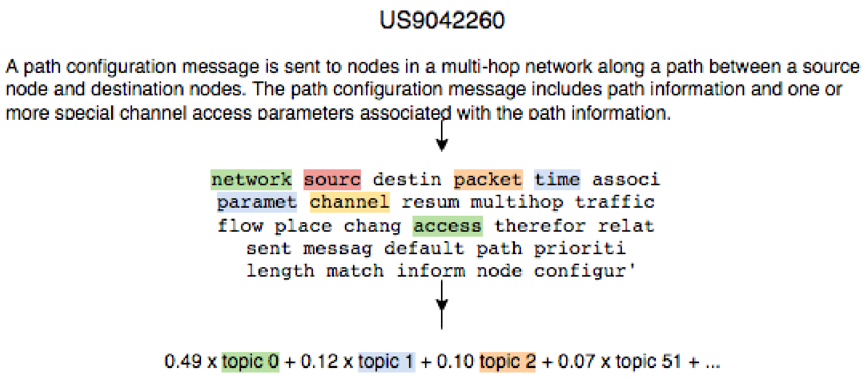


Figure A.3: Example of a patent converted into a distribution over topics.

B. Citations and Patent Vector Similarity

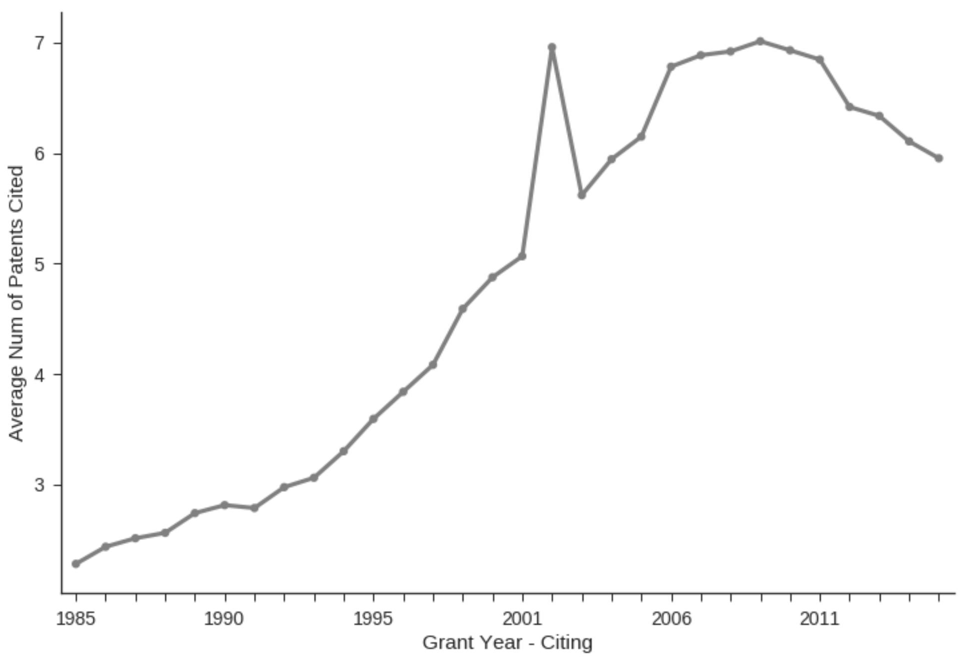


Figure B.1: Averate number of patents cited over time

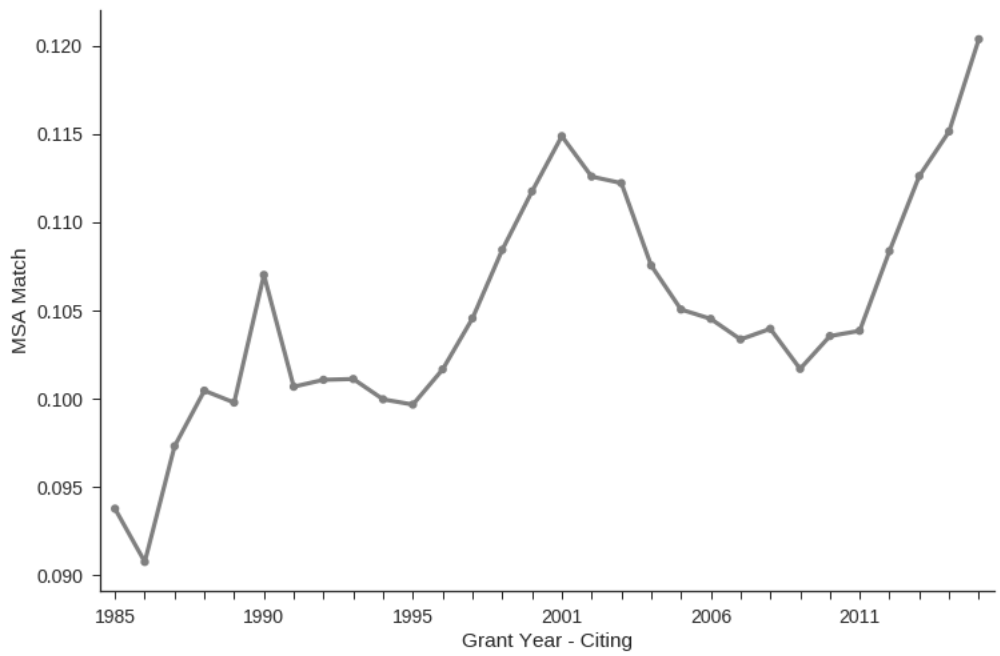


Figure B.2: Proportion of cited patents in the same MSA over time

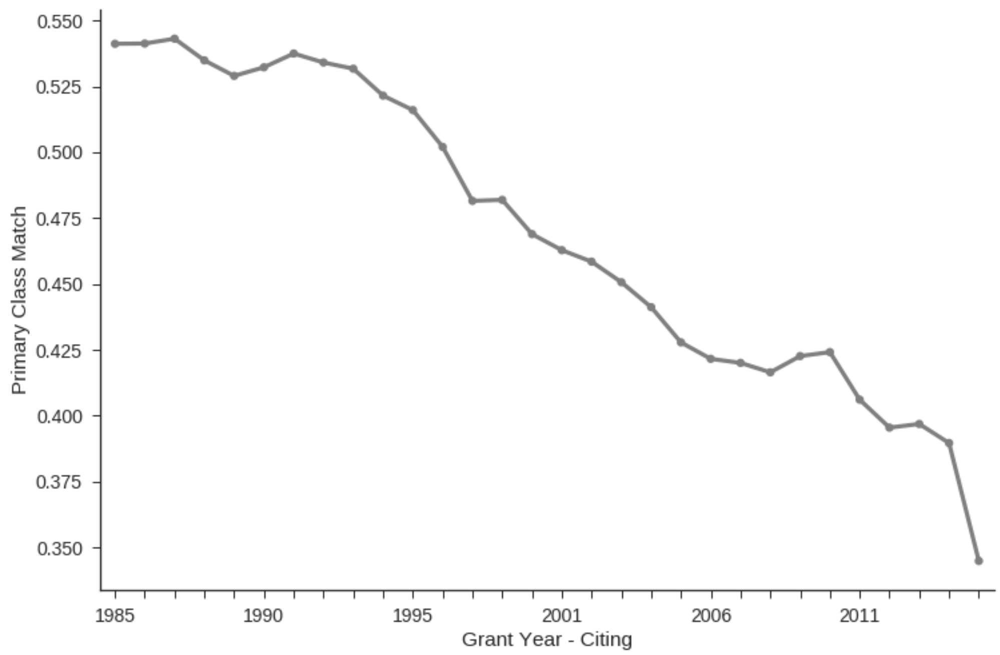


Figure B.3: Proportion of cited patents in the same primary class over time

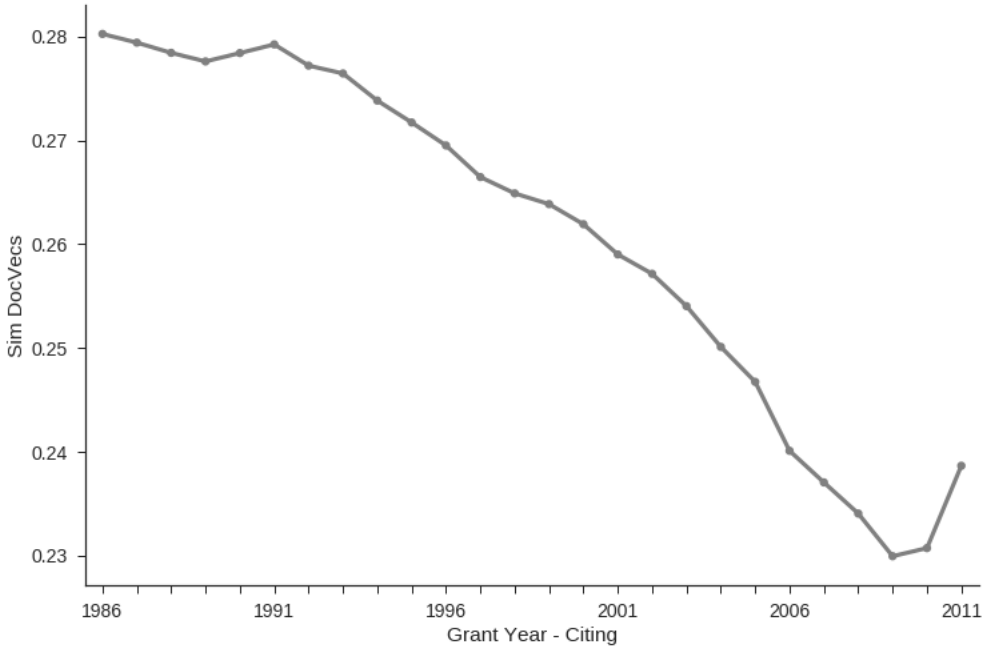


Figure B.4: Average DocVecs similarity to cited patents

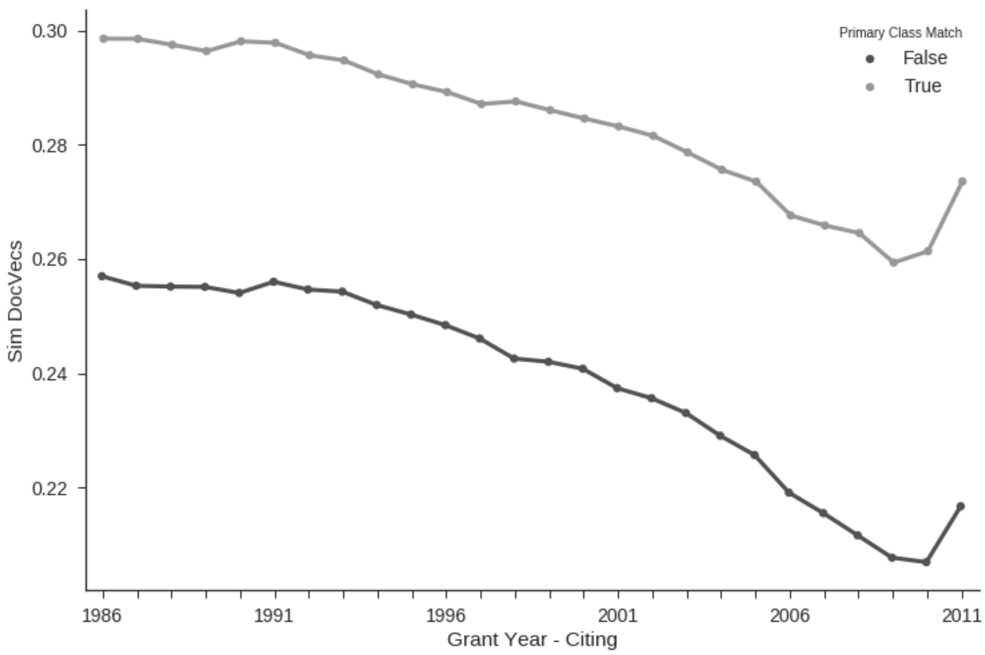


Figure B.5: Average DocVecs similarity to cited patents in the same primary class

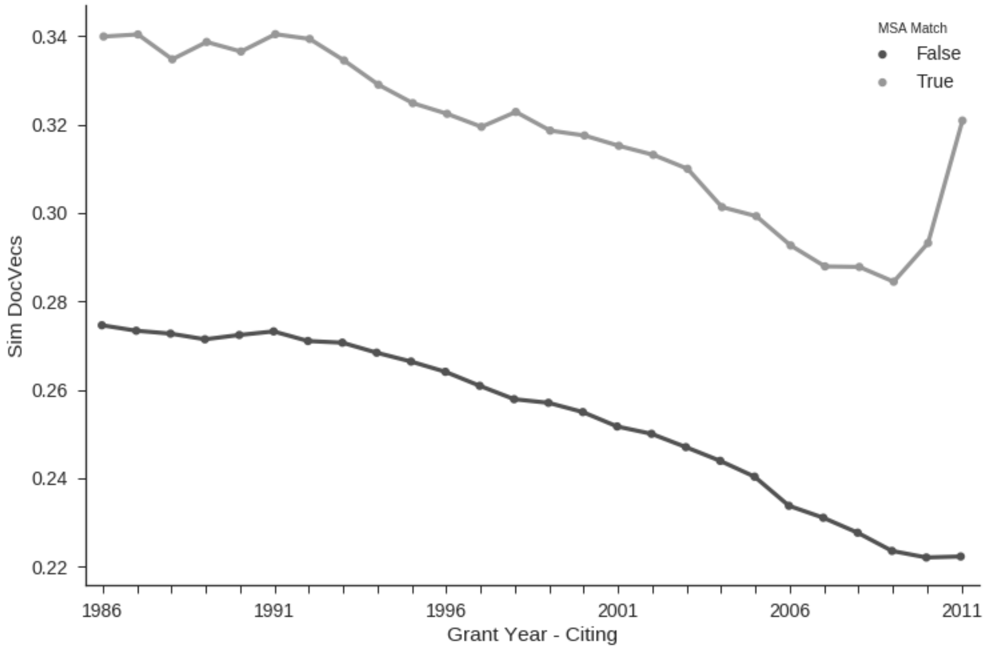


Figure B.6: Average DocVecs similarity to cited patents in the same MSA

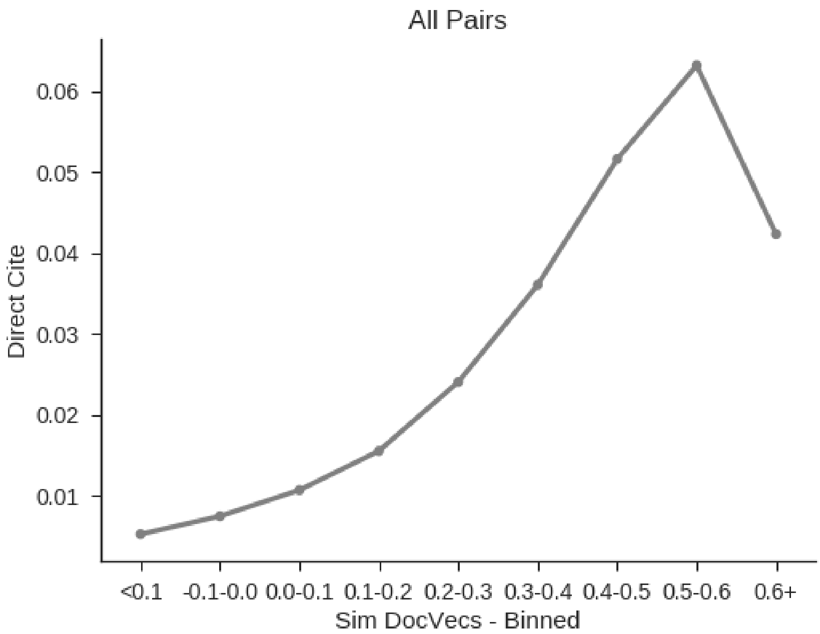


Figure B.7: Rate of direct citation conditional on level of DocVecs Similarity, All Pairs

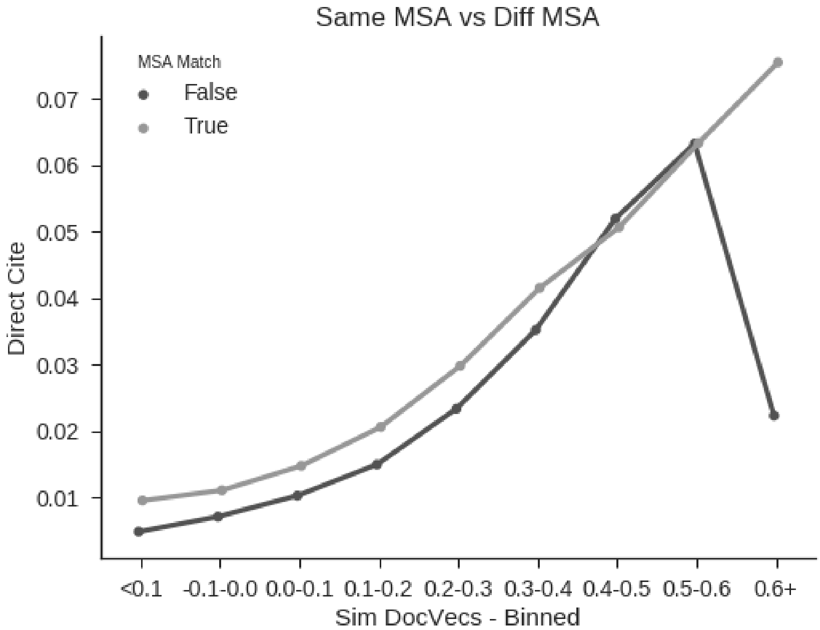


Figure B.8: Rate of direct citation conditional on level of DocVecs Similarity, Same MSA vs Different MSA

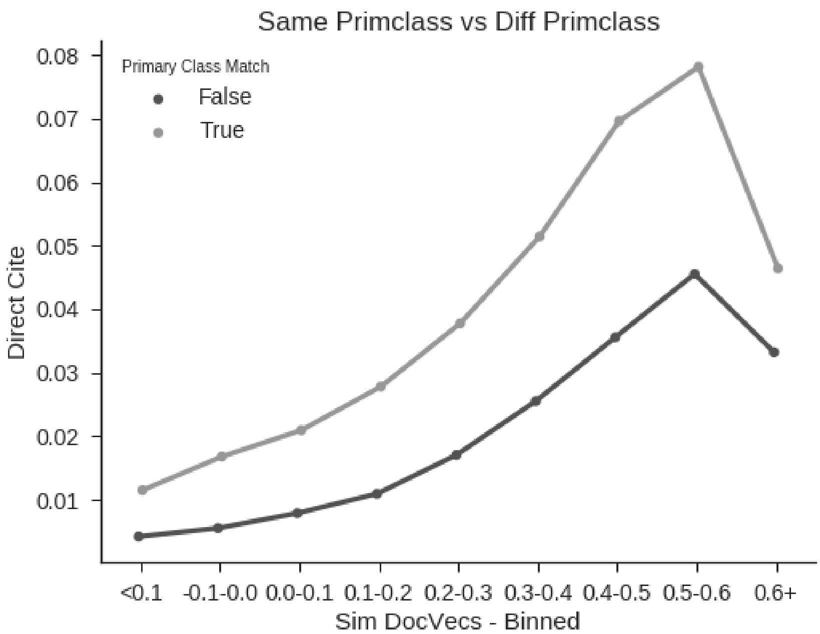


Figure B.9: Rate of direct citation conditional on level of DocVecs Similarity, Same Primary Class vs Different Primary Class

	<0.1	-0.1-0.0	0.0-0.1	0.1-0.2	0.2-0.3	0.3-0.4	0.4-0.5	0.5-0.6	0.6+
All Pairs, N	51355	205500	498927	685041	562529	293309	102019	27531	18958
All Pairs, Prop Cited	0.005	0.007	0.011	0.016	0.024	0.036	0.052	0.063	0.042
Same MSA, N	3768	16056	42380	65643	63246	41573	19994	8327	7163
Same MSA, Prop Cited	0.01	0.011	0.015	0.021	0.03	0.041	0.051	0.063	0.075
Diff MSA, N	47587	189444	456547	619398	499283	251736	82025	19204	11795
Diff MSA, Prop Cited	0.005	0.007	0.01	0.015	0.023	0.035	0.052	0.063	0.022
p-value	0	0	0	0	0	0	0.466	0.982	0
Same NAICS, N	21756	92343	239948	354306	313789	175065	64802	18362	13949
Same NAICS, Prop Cited	0.006	0.009	0.013	0.019	0.028	0.042	0.059	0.072	0.047
Diff NAICS, N	29599	113157	258979	330735	248740	118244	37217	9169	5009
Diff NAICS, Prop Cited	0.005	0.006	0.009	0.012	0.019	0.028	0.039	0.046	0.029
p-value	0.355	0	0	0	0	0	0	0	0
Same Primclass, N	7035	34445	105794	185777	190332	119502	48211	14968	13148
Same Primclass, Prop Cited	0.012	0.017	0.021	0.028	0.038	0.051	0.07	0.078	0.046
Diff Primclass, N	44320	171055	393133	499264	372197	173807	53808	12563	5810
Diff Primclass, Prop Cited	0.004	0.006	0.008	0.011	0.017	0.026	0.036	0.046	0.033
p-value	0	0	0	0	0	0	0	0	0

Table B.1: Summary table of rates of direct citation by DocVecs similarity

C. Inventor Mobility

	<0.1	-0.1-0.0	0.0-0.1	0.1-0.2	0.2-0.3	0.3-0.4	0.4-0.5	0.5-0.6	0.6+
Before Firm Change, N	3437	12802	32168	49481	50733	36190	19608	10204	9301
Before Firm Change, Prop Cites	0.066	0.072	0.08	0.089	0.11	0.142	0.199	0.265	0.28
After Firm Change, N	3018	8594	18736	24880	22741	15429	7683	2702	1452
After Firm Change, Prop Cites	0.01	0.012	0.022	0.039	0.062	0.091	0.116	0.174	0.294
Diff, p-value	0	0	0	0	0	0	0	0	0.278

Table C.1: Rate of self-citation before and after firm change

	<0.1	-0.1-0.0	0.0-0.1	0.1-0.2	0.2-0.3	0.3-0.4	0.4-0.5	0.5-0.6	0.6+
Before Firm Change, N	3620	12775	32021	49163	50138	35447	18975	9756	9033
Before Firm Change, Pct Common Cites	0.039	0.034	0.044	0.062	0.086	0.135	0.235	0.419	0.658
After Firm Change, N	3205	8593	18731	24868	22734	15419	7678	2701	1452
After Firm Change, Pct Common Cites	0.007	0.009	0.015	0.028	0.046	0.071	0.107	0.211	0.52
Diff, p-value	0	0	0	0	0	0	0	0	0

Table C.2: Rate of self-citation before and after firm change

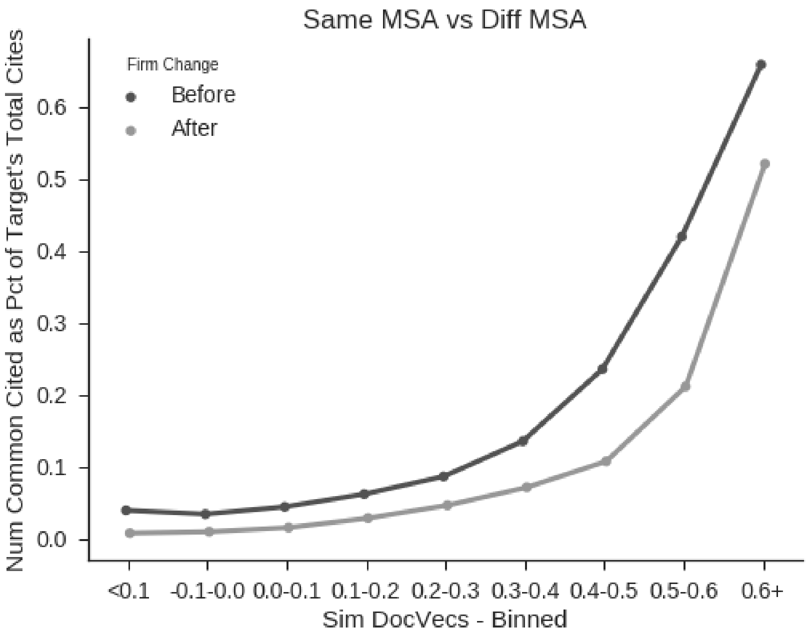


Figure C.1: Rate of direct citation conditional on level of DocVecs Similarity, Same Primary Class vs Different Primary Class

D. Regressions

KS = Pct Cites in Target's MSA, JTH Sample			
	1975-85	1985-95	1995-05
$I(MSAMatch)$	0.0523***	0.0616***	0.0642***
	(0.0016)	(0.0011)	(0.0009)
N	58647	107358	185154
Adjusted R^2	0.03	0.05	0.05
Year FE	True	True	True
PC FE	True	True	True

KS = Similarity to Citations, JTH Sample			
	1975-85	1985-95	1995-05
$I(MSAMatch)$	0.0540***	0.0469***	0.0422***
	(0.0021)	(0.0014)	(0.0010)
N	38541	69612	122217
Adjusted R^2	0.05	0.05	0.06
Year FE	True	True	True
PC FE	True	True	True

Table D.1: Baseline regression results for JTH Sample, Raw data.

$KS =$ Sim DocVecs, Within-NAICS				
	1975-85	1985-95	1995-05	2005-15
$I(MSAMatch)$	0.0044***	0.0082***	0.0082***	0.0077***
	(0.0007)	(0.0006)	(0.0005)	(0.0004)
N	192841	281222	437685	569252
Adjusted R^2	0.07	0.07	0.08	0.06
Year FE	True	True	True	True
PC FE	True	True	True	True

$KS =$ Sim DocVecs, Within-Primclass				
	1975-85	1985-95	1995-05	2005-15
$I(MSAMatch)$	0.0073***	0.0111***	0.0109***	0.0079***
	(0.0009)	(0.0007)	(0.0005)	(0.0004)
N	170882	252174	401623	529686
Adjusted R^2	0.07	0.07	0.08	0.07
Year FE	True	True	True	True
PC FE	True	True	True	True

$KS =$ Pct Common Cited, Within-NAICS				
	1975-85	1985-95	1995-05	2005-15
$I(MSAMatch)$	0.0005***	0.0012***	0.0013***	0.0011***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
N	188734	273714	384782	552981
Adjusted R^2	0.00	0.00	0.00	0.00
Year FE	True	True	True	True
PC FE	True	True	True	True

$KS =$ Pct Common Cited, Within-Primclass				
	1975-85	1985-95	1995-05	2005-15
$I(MSAMatch)$	0.0038***	0.0056***	0.0049***	0.0044***
	(0.0003)	(0.0003)	(0.0002)	(0.0002)
N	167005	245355	352428	514322
Adjusted R^2	0.01	0.01	0.01	0.01
Year FE	True	True	True	True
PC FE	True	True	True	True

Table D.2: Baseline regression results for Number of Common Cited and Sim DocVecs, Raw data.

Inventor Match				
	1975-85	1985-95	1995-05	2005-15
$I(InvMatch)$	1.2789***	1.5206***	1.3817***	1.2664***
	(0.1042)	(0.0713)	(0.0559)	(0.0563)
N	192841	281222	437685	569252
Adjusted R^2	0.07	0.07	0.08	0.06
Year FE	True	True	True	True
PC FE	True	True	True	True

Common Cited Patent ≥ 1				
	1975-85	1985-95	1995-05	2005-15
$I(CommonCited \geq 1)$	1.5084***	1.2870***	1.1149***	0.8388***
	(0.0963)	(0.0637)	(0.0398)	(0.0254)
N	192841	281222	437685	569252
Adjusted R^2	0.07	0.07	0.08	0.06
Year FE	True	True	True	True
PC FE	True	True	True	True

Lawyer Match				
	1975-85	1985-95	1995-05	2005-15
$I(lawyer_i = lawyer_j)$	0.2840***	0.3809***	0.4535***	0.3772***
	(0.0364)	(0.0281)	(0.0299)	(0.0275)
N	192841	281222	437685	569252
Adjusted R^2	0.07	0.07	0.08	0.06
Year FE	True	True	True	True
PC FE	True	True	True	True

Primary Class Match				
	1975-85	1985-95	1995-05	2005-15
$I(PrimclassMatch)$	0.4413***	0.4449***	0.4291***	0.4045***
	(0.0085)	(0.0066)	(0.0050)	(0.0042)
N	192841	281222	437685	569252
Adjusted R^2	0.08	0.09	0.10	0.07
Year FE	True	True	True	True
PC FE	True	True	True	True

Table D.3: Regression results for other match variables, DocVecs Similarity.

Inventor Match				
	1975-85	1985-95	1995-05	2005-15
$I(InvMatch)$	6.4372***	11.3071***	11.8043***	13.5321***
	(1.2517)	(1.1416)	(1.0102)	(1.0385)
N	188734	273714	384782	552981
Adjusted R^2	0.08	0.15	0.15	0.16
Year FE	True	True	True	True
PC FE	True	True	True	True

Common Cited Patent ≥ 1				
	1975-85	1985-95	1995-05	2005-15
$I(CommonCited \geq 1)$	16.9106***	18.1456***	14.2776***	8.9272***
	(1.2517)	(0.9336)	(0.6112)	(0.3469)
N	188734	273714	384782	552981
Adjusted R^2	0.55	0.48	0.39	0.26
Year FE	True	True	True	True
PC FE	True	True	True	True

Lawyer Match				
	1975-85	1985-95	1995-05	2005-15
$I(lawyer_i = lawyer_j)$	0.8125***	1.8604***	2.8891***	2.6163***
	(0.2014)	(0.2385)	(0.3244)	(0.2857)
N	188734	273714	384782	552981
Adjusted R^2	0.01	0.02	0.03	0.02
Year FE	True	True	True	True
PC FE	True	True	True	True

Primary Class Match				
	1975-85	1985-95	1995-05	2005-15
$I(PrimclassMatch)$	0.1246***	0.2104***	0.2154***	0.1843***
	(0.0156)	(0.0178)	(0.0140)	(0.0102)
N	188734	273714	384782	552981
Adjusted R^2	0.00	0.01	0.00	0.01
Year FE	True	True	True	True
PC FE	True	True	True	True

Table D.4: Regression results for other match variables, Proportion of Common Cited patents.

Inventor Match				
	1975-85	1985-95	1995-05	2005-15
$I(InvMatch)$	0.1727***	0.2053***	0.1865***	0.1710***
	(0.0141)	(0.0096)	(0.0076)	(0.0076)
N	192841	281222	437685	569252
Adjusted R^2	0.07	0.07	0.08	0.06
Year FE	True	True	True	True
PC FE	True	True	True	True

Common Cited Patent ≥ 1				
	1975-85	1985-95	1995-05	2005-15
$I(CommonCited \geq 1)$	0.2037***	0.1738***	0.1505***	0.1132***
	(0.0130)	(0.0086)	(0.0054)	(0.0034)
N	192841	281222	437685	569252
Adjusted R^2	0.07	0.07	0.08	0.06
Year FE	True	True	True	True
PC FE	True	True	True	True

Lawyer Match				
	1975-85	1985-95	1995-05	2005-15
$I(lawyer_i = lawyer_j)$	0.0383***	0.0514***	0.0612***	0.0509***
	(0.0049)	(0.0038)	(0.0040)	(0.0037)
N	192841	281222	437685	569252
Adjusted R^2	0.07	0.07	0.08	0.06
Year FE	True	True	True	True
PC FE	True	True	True	True

Primary Class Match				
	1975-85	1985-95	1995-05	2005-15
$I(PrimclassMatch)$	0.0596***	0.0601***	0.0579***	0.0546***
	(0.0012)	(0.0009)	(0.0007)	(0.0006)
N	192841	281222	437685	569252
Adjusted R^2	0.08	0.09	0.10	0.07
Year FE	True	True	True	True
PC FE	True	True	True	True

Table D.5: Regression results for other match variables, DocVecs Similarity, Raw Data.

	Inventor Match			
	1975-85	1985-95	1995-05	2005-15
$I(InvMatch)$	0.0966***	0.1697***	0.1772***	0.2031***
	(0.0188)	(0.0171)	(0.0152)	(0.0156)
N	188734	273714	384782	552981
Adjusted R^2	0.08	0.15	0.15	0.16
Year FE	True	True	True	True
PC FE	True	True	True	True
	Common Cited Patent ≥ 1			
	1975-85	1985-95	1995-05	2005-15
$I(CommonCited \geq 1)$	0.2539***	0.2724***	0.2143***	0.1340***
	(0.0188)	(0.0140)	(0.0092)	(0.0052)
N	188734	273714	384782	552981
Adjusted R^2	0.55	0.48	0.39	0.26
Year FE	True	True	True	True
PC FE	True	True	True	True
	Lawyer Match			
	1975-85	1985-95	1995-05	2005-15
$I(lawyer_i = lawyer_j)$	0.0122***	0.0279***	0.0434***	0.0393***
	(0.0030)	(0.0036)	(0.0049)	(0.0043)
N	188734	273714	384782	552981
Adjusted R^2	0.01	0.02	0.03	0.02
Year FE	True	True	True	True
PC FE	True	True	True	True
	Primary Class Match			
	1975-85	1985-95	1995-05	2005-15
$I(PrimclassMatch)$	0.0019***	0.0032***	0.0032***	0.0028***
	(0.0002)	(0.0003)	(0.0002)	(0.0002)
N	188734	273714	384782	552981
Adjusted R^2	0.00	0.01	0.00	0.01
Year FE	True	True	True	True
PC FE	True	True	True	True

Table D.6: Regression results for other match variables, Proportion of Common Cited patents, Raw data.

NAICS Sample: Primclass Similarity								
	(1)				(2)			
$KS = sim_{DV}(i, j)$	1980-85	1985-95	1995-05	2005-15	1980-85	1985-95	1995-05	2005-15
$I(MSAMatch)$	0.0023	0.0053***	0.0041***	0.0037***	0.0028	0.0072***	0.0020	-0.0000
	(0.0016)	(0.0009)	(0.0006)	(0.0005)	(0.0049)	(0.0027)	(0.0019)	(0.0020)
$I_{MSA} * sim_{DV}(pc_i, pc_j)$					-0.0038	-0.0141	0.0158	0.0270*
					(0.0367)	(0.0211)	(0.0149)	(0.0144)
$I(PrimclassMatch)$	0.0018	0.0037**	0.0032***	0.0019**	0.0018	0.0037**	0.0031***	0.0018*
	(0.0030)	(0.0017)	(0.0012)	(0.0009)	(0.0030)	(0.0017)	(0.0012)	(0.0009)
$sim_{DV}(pc_i, pc_j)$	0.8893***	0.9236***	0.8929***	0.9294***	0.8902***	0.9273***	0.8888***	0.9220***
	(0.0286)	(0.0173)	(0.0126)	(0.0116)	(0.0301)	(0.0182)	(0.0132)	(0.0122)
N	40323	110982	215861	344313	40323	110982	215861	344313
Adjusted R^2	0.12	0.13	0.13	0.08	0.12	0.13	0.13	0.08
Year FE	True	True	True	True	True	True	True	True
PC FE	True	True	True	True	True	True	True	True
Inv & Lawyer Match	True	True	True	True	True	True	True	True

NAICS Sample: Primclass-MSA Similarity								
	(3)				(4)			
$KS = sim_{DV}(i, j)$	1980-85	1985-95	1995-05	2005-15	1980-85	1985-95	1995-05	2005-15
$I(MSAMatch)$	0.0015	0.0048***	0.0025***	0.0023***	-0.0002	0.0038*	-0.0000	-0.0043***
	(0.0016)	(0.0010)	(0.0007)	(0.0005)	(0.0035)	(0.0020)	(0.0015)	(0.0015)
$I_{MSA} * sim_{DV}(pc_{i,MSA_i}, pc_{j,MSA_j})$					0.0113	0.0070	0.0181*	0.0461***
					(0.0241)	(0.0139)	(0.0108)	(0.0101)
$I(PrimclassMatch)$	0.0389***	0.0415***	0.0355***	0.0320***	0.0388***	0.0414***	0.0353***	0.0317***
	(0.0026)	(0.0015)	(0.0010)	(0.0008)	(0.0026)	(0.0015)	(0.0010)	(0.0008)
$sim_{DV}(pc_{i,MSA_i}, pc_{j,MSA_j})$	0.2227***	0.2227***	0.2907***	0.3238***	0.2203***	0.2211***	0.2863***	0.3122***
	(0.0123)	(0.0073)	(0.0061)	(0.0055)	(0.0135)	(0.0080)	(0.0067)	(0.0061)
N	40323	110982	215861	344313	40323	110982	215861	344313
Adjusted R^2	0.11	0.11	0.12	0.08	0.11	0.11	0.12	0.08
Year FE	True	True	True	True	True	True	True	True
PC FE	True	True	True	True	True	True	True	True
Inv & Lawyer Match	True	True	True	True	True	True	True	True

Table D.7: Regression results including technology proximity and other controls with raw data for DocVec similarity.

Primclass Sample: Primclass-MSA Similarity								
	(5)				(6)			
$KS = sim_{DV}(i,j)$	1980-85	1985-95	1995-05	2005-15	1980-85	1985-95	1995-05	2005-15
$I(MSAMatch)$	0.0014	0.0034***	0.0044***	0.0022***	-0.0046	0.0081**	0.0091***	0.0054**
	(0.0020)	(0.0011)	(0.0008)	(0.0006)	(0.0072)	(0.0039)	(0.0029)	(0.0025)
$I_{MSA} * sim_{DV}(pc_{i,MSA_i}, pc_{j,MSA_j})$					0.0264	-0.0210	-0.0224	-0.0158
					(0.0317)	(0.0179)	(0.0140)	(0.0123)
$sim_{DV}(pc_{i,MSA_i}, pc_{j,MSA_j})$	0.1188***	0.1183***	0.1500***	0.1698***	0.1148***	0.1218***	0.1542***	0.1730***
	(0.0136)	(0.0082)	(0.0069)	(0.0064)	(0.0144)	(0.0087)	(0.0075)	(0.0069)
N	38324	106152	205248	324142	38324	106152	205248	324142
Adjusted R^2	0.08	0.08	0.09	0.06	0.08	0.08	0.09	0.06
Year FE	True	True	True	True	True	True	True	True
PC FE	True	True	True	True	True	True	True	True
Inv & Lawyer Match	True	True	True	True	True	True	True	True

Table D.8: Regression results including technology proximity and other controls with raw data for DocVec similarity.

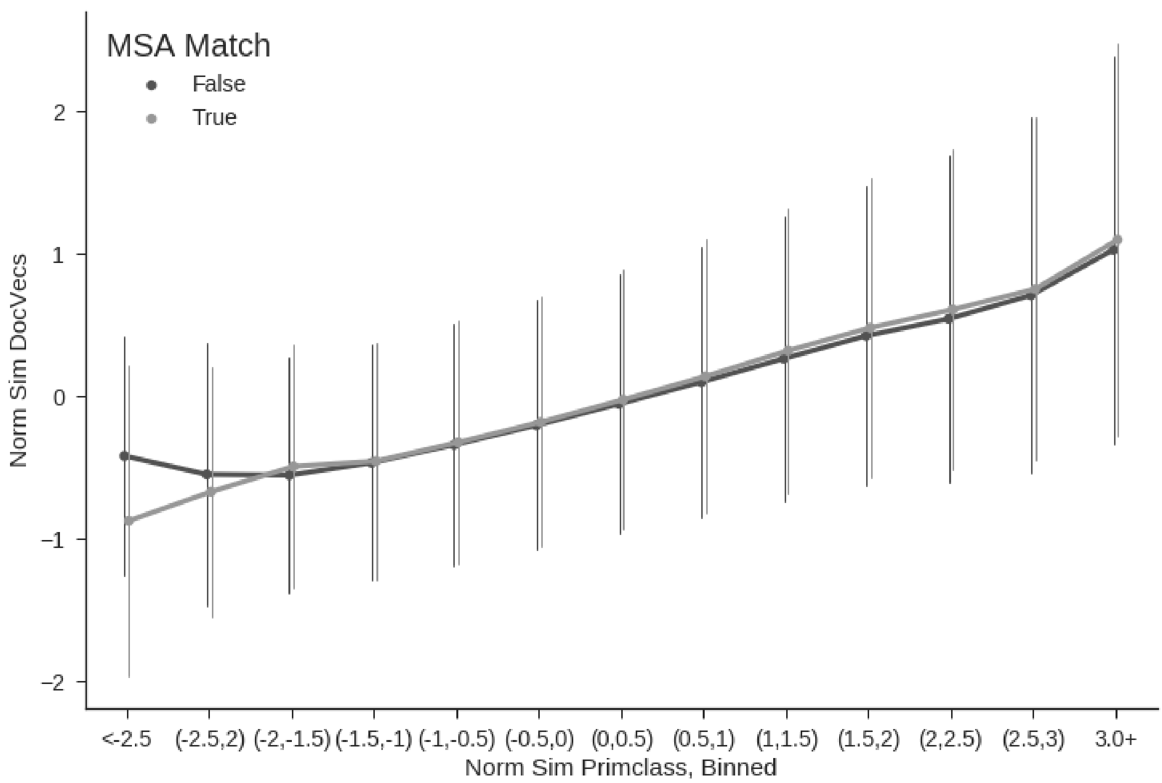


Figure D.1: Conditional means of DocVec similarity by primary class similarity, within-NAICS sample.

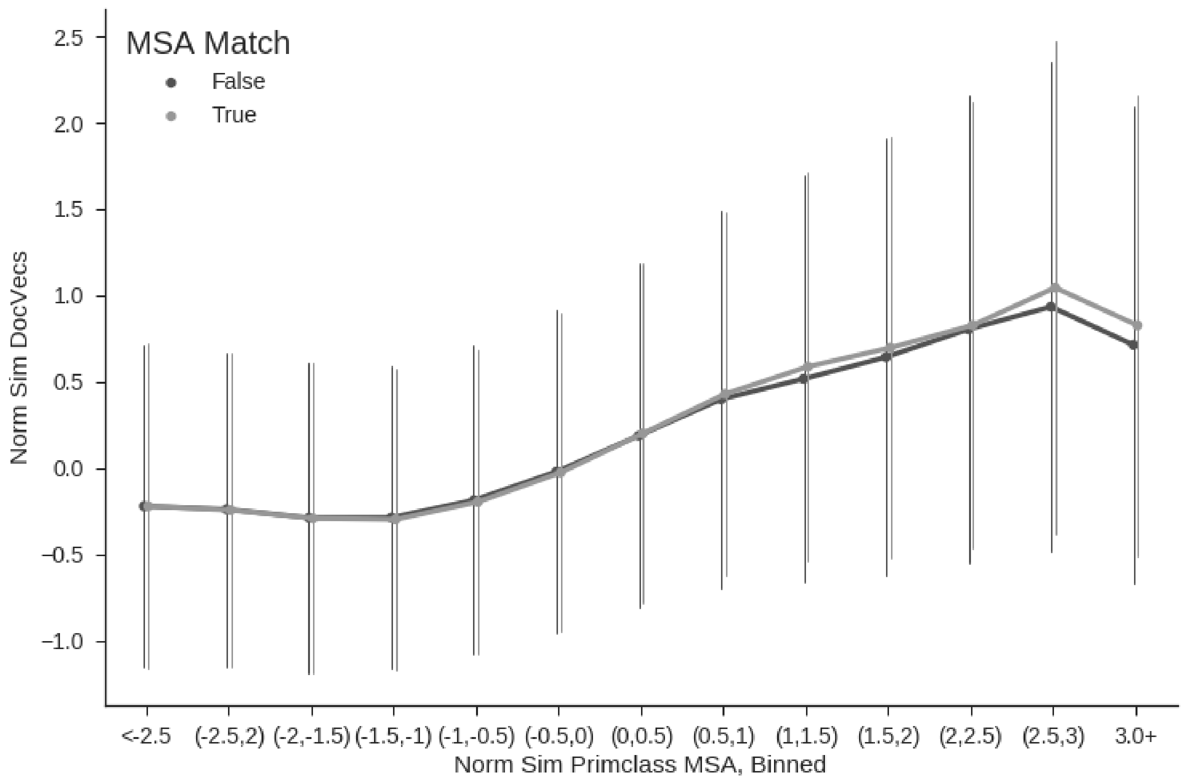


Figure D.2: Conditional means of DocVec similarity by MSA-primary class similarity, within-NAICS sample.

	<-2.5	(-2.5,-2)	(-2,-1.5)	(-1.5,-1)	(-1,-0.5)	(-0.5,0)	(0,0.5)	(0.5,1)	(1,1.5)	(1.5,2)	(2,2.5)	(2.5,3)	3<
Norm Sim DocVecs, MSA Match = T, Mean	-0.192	-0.095	-0.071	-0.078	-0.257	-0.314	-0.142	0.046	0.187	0.302	0.299	0.286	0.265
Norm Sim DocVecs, MSA Match = F, Mean	-0.172	-0.186	-0.186	-0.219	-0.288	-0.248	-0.101	0.078	0.21	0.28	0.293	0.278	0.393
MSA Match = T, N	126	115	262	454	1550	16859	73790	79223	46695	19026	7292	3738	6496
MSA Match = F, N	2318	2465	5417	13440	45020	151021	278309	195291	89770	32034	12363	5630	5970
Diff in Mean	-0.02	0.091	0.115	0.141	0.031	-0.067	-0.041	-0.032	-0.022	0.022	0.006	0.008	-0.127
t-value	-0.232	0.983	1.833	3.069	1.275	-9.044	-10.734	-7.951	-3.862	2.274	0.365	0.35	-6.045
p-value	0.817	0.327	0.068	0.002	0.202	0	0	0	0	0.023	0.715	0.727	0

Table D.9: DocVec similarity conditional on primary class-MSA similarity for within-primary class patent pairs. Pharmaceutical drugs primary class 514 is omitted.

	1980-85	1985-95	1995-05	2005-15
$I(MSAMatch)$	-0.0296	0.0305**	0.0276***	0.0097
	(0.0241)	(0.0123)	(0.0083)	(0.0065)
$I_{MSA} * sim_{DV}(pc_{i,MSA_i}, pc_{j,MSA_j})$	0.0412	-0.0146	0.0021	0.0064
	(0.0266)	(0.0144)	(0.0106)	(0.0092)
$I_{MSA} * sim_{DV}(pc_{i,MSA_i}, pc_{j,MSA_j})^3$	-0.0018	0.0015	-0.0007	-0.0010
	(0.0028)	(0.0016)	(0.0012)	(0.0011)
$sim_{DV}(pc_{i,MSA_i}, pc_{j,MSA_j})$	0.0826***	0.0851***	0.1030***	0.1121***
	(0.0108)	(0.0065)	(0.0053)	(0.0049)
$sim_{DV}(pc_{i,MSA_i}, pc_{j,MSA_j})^3$	-0.0030**	-0.0030***	-0.0032***	-0.0029***
	(0.0014)	(0.0009)	(0.0008)	(0.0007)
N	38324	106152	205248	324142
Adjusted R^2	0.08	0.08	0.09	0.06
Year FE	True	True	True	True
PC FE	True	True	True	True
Inv & Lawyer Match	True	True	True	True

Table D.10: Regression results with primary class-MSA similarity with cubic polynomial controls for within-primary class patent pairs.

E. New Terms

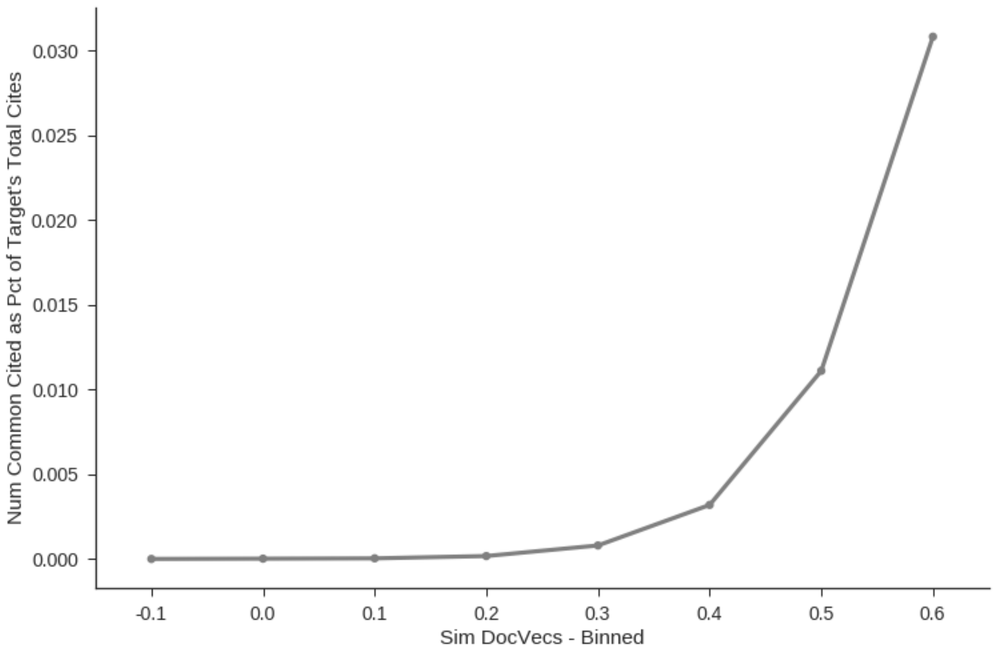


Figure E.1: Number of common cited patents as a percentage of target's total citations, by DocVecs similarity

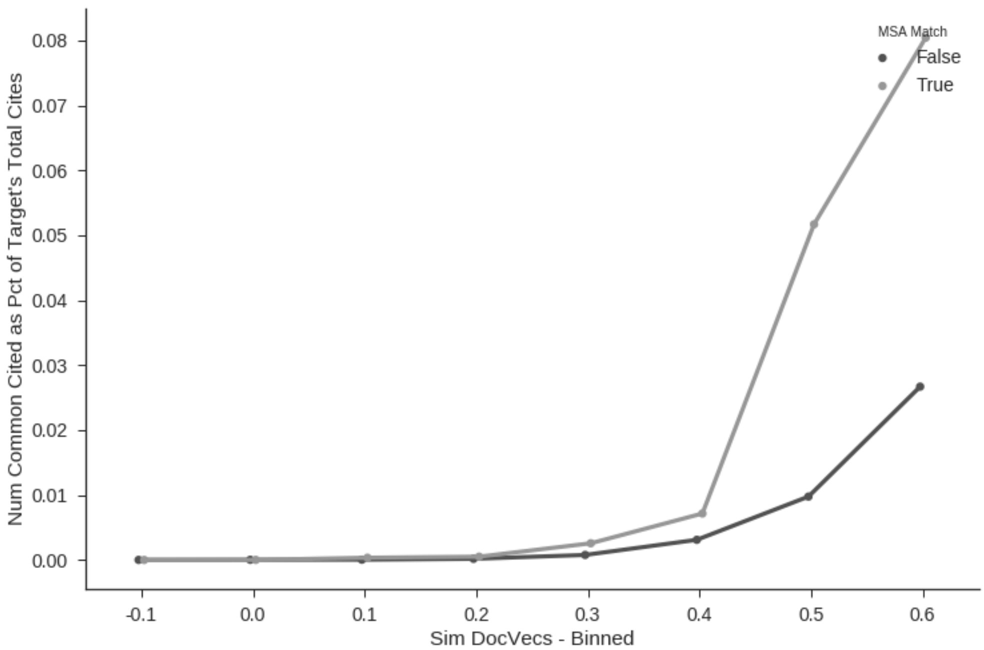
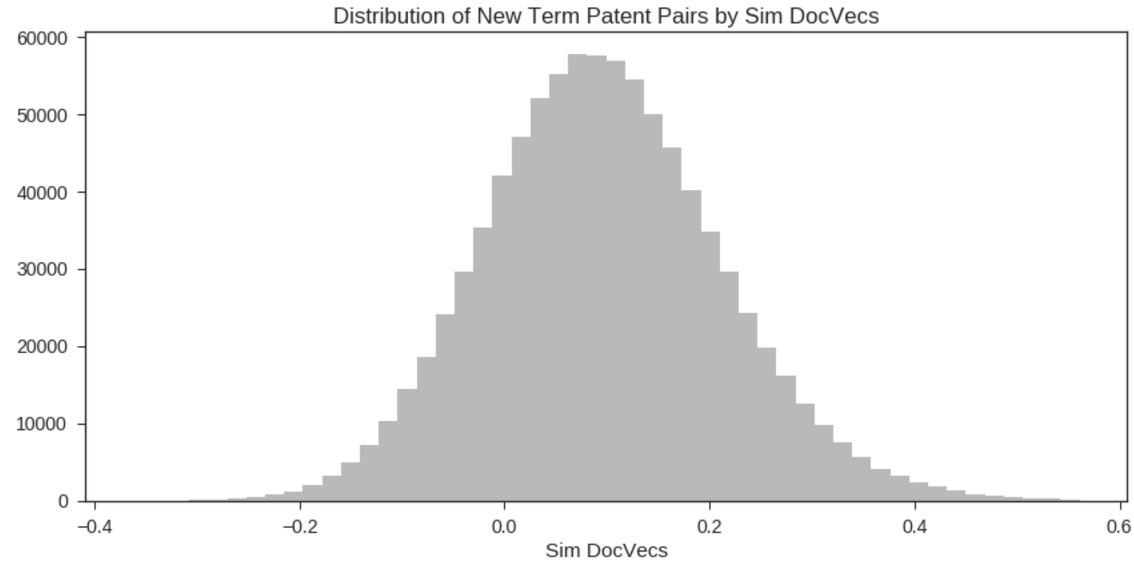
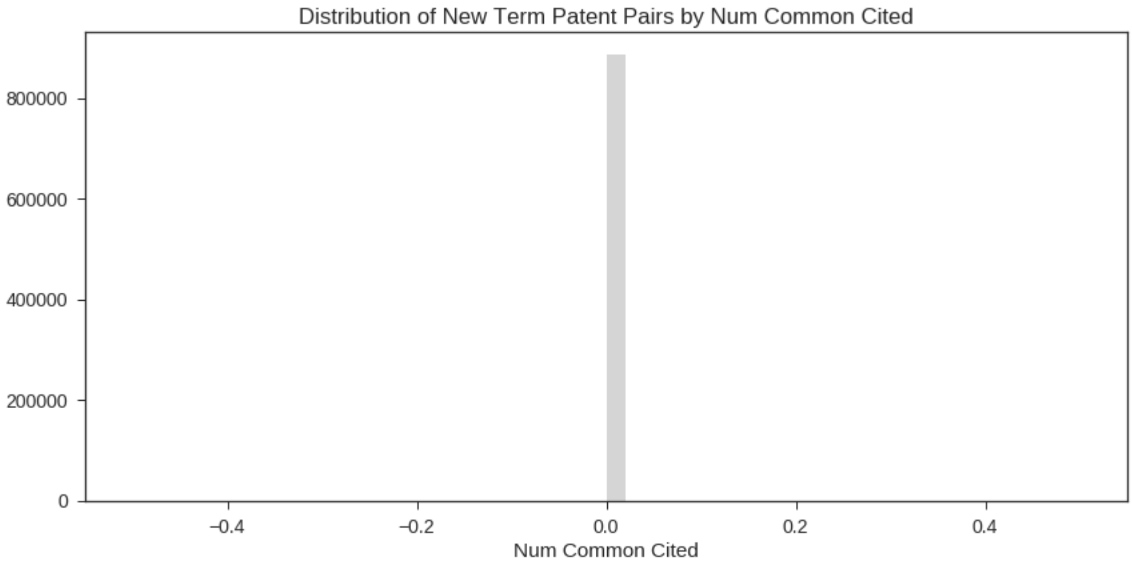


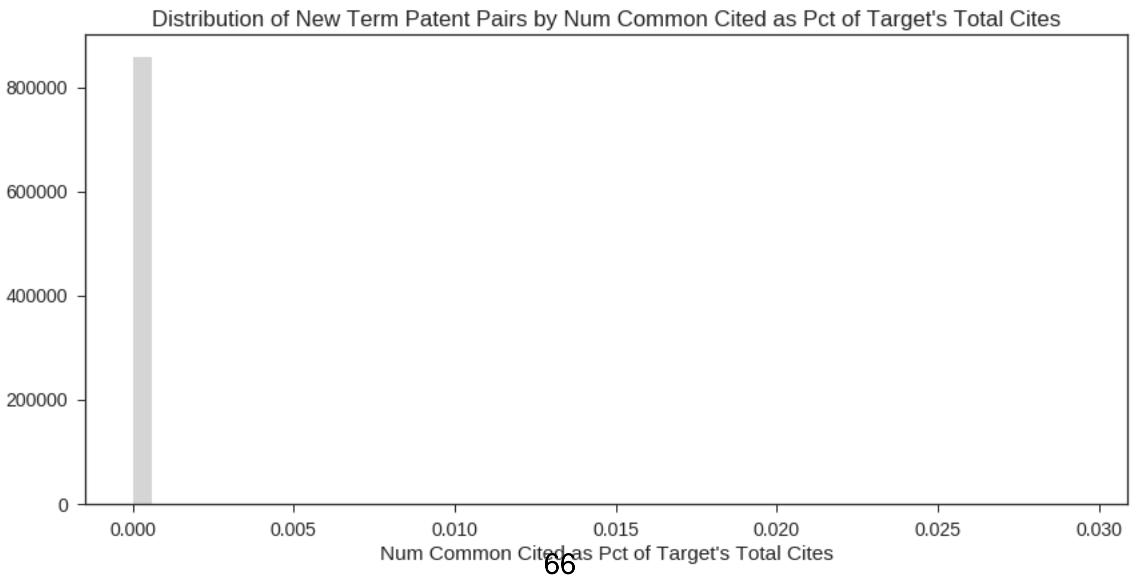
Figure E.2: Number of common cited patents as a percentage of target's total citations, by DocVecs similarity

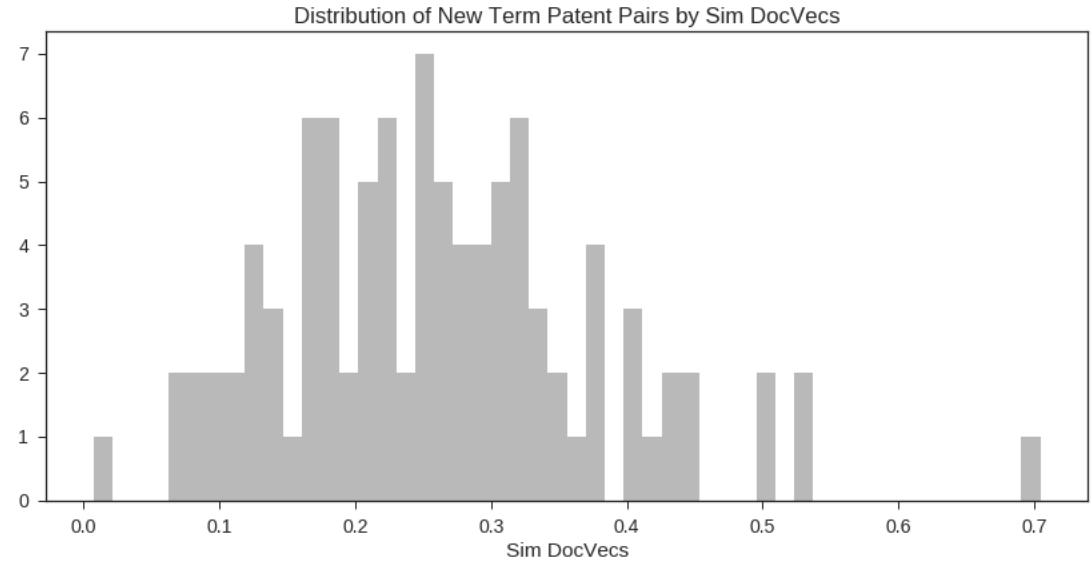
	Sim DocVecs	Num Common Cited	Pct Common Cited, Target's Citations	First Year	First Year, Num Pats	First Year, Num Pairs
gui	0.09	0.00	0.00	1992	1348	796727
lun	0.12	0.00	0.00	1995	415	63044
asic	0.11	0.00	0.00	1987	198	16716
url	0.10	0.01	0.00	1995	111	4847
serd	0.12	0.02	0.00	1998	75	1929
chat	0.10	0.00	0.00	1992	9	1299
bist	0.12	0.00	0.00	1990	42	810
femto	0.15	0.04	0.00	2007	27	563
angst	0.16	0.00	0.00	1994	40	549
mcm	0.10	0.00	0.00	1991	32	440
www	0.15	0.01	0.01	1995	9	291
efus	0.12	0.02	0.00	2000	22	201
femtocel	0.22	0.09	0.00	2007	8	198
adenovir	0.26	0.03	0.00	1993	15	98
cyclin	0.18	0.00	0.00	1991	13	80
n 1	0.13	0.00	0.00	1998	13	63
dvd	0.16	0.00	0.00	1996	10	44
websit	0.18	0.00	0.00	1996	10	42
gpu	0.05	0.00	0.00	2001	9	36
pcie	0.21	0.00	0.00	2004	9	35

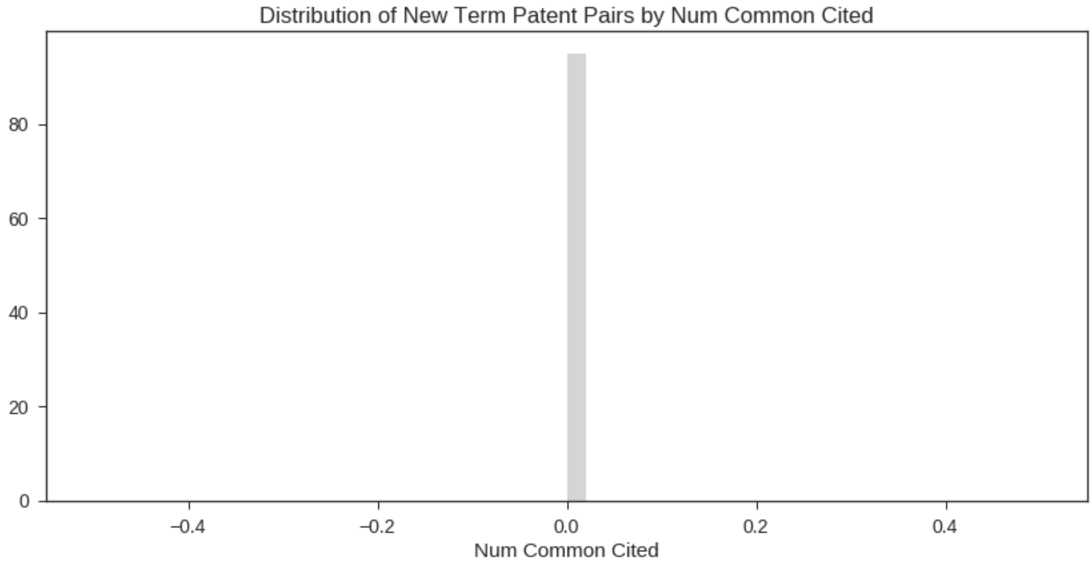
Table E.1: Comparison of similarity and backward citation overlap for new patents using new terms.

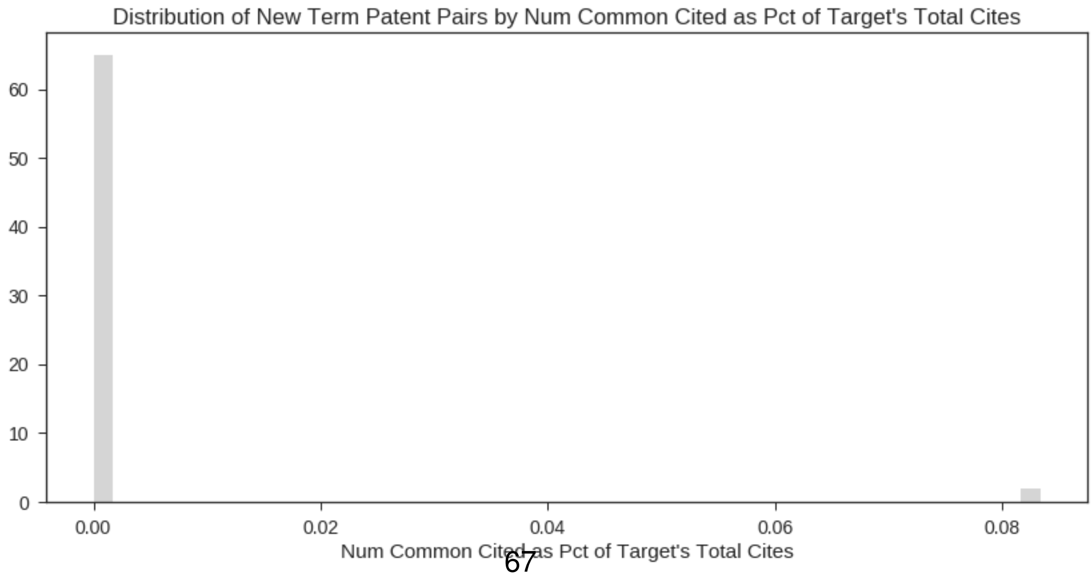

Distribution of New Term Patent Pairs by Sim DocVecs
60000
50000
40000
30000
20000
10000
0
−0.4
−0.2
0.0
0.2
0.4
0.6
Sim DocVecs


Distribution of New Term Patent Pairs by Num Common Cited
800000
600000
400000
200000
0
−0.4
−0.2
0.0
0.2
0.4
Num Common Cited


Distribution of New Term Patent Pairs by Num Common Cited as Pct of Target's Total Cites
800000
600000
400000
200000
0
0.000
0.005
0.010
0.015
0.020
0.025
0.030
Num Common Cited as Pct of Target's Total Cites


Distribution of New Term Patent Pairs by Sim DocVecs
0
1
2
3
4
5
6
7
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
Sim DocVecs


Distribution of New Term Patent Pairs by Num Common Cited
0
20
40
60
80
−0.4
−0.2
0.0
0.2
0.4
Num Common Cited


Distribution of New Term Patent Pairs by Num Common Cited as Pct of Target's Total Cites
0
10
20
30
40
50
60
0.00
0.02
0.04
0.06
0.08
Num Common Cited as Pct of Target's Total Cites

References

Ajay Agrawal, Iain Cockburn, and John McHale. Gone but not forgotten: knowledge flows, labor mobility, and enduring social relationships. *Journal of Economic Geography*, 6(5): 571–591, 2006.

Juan Alcacer and Michelle Gittelman. Patent citations as a measure of knowledge flows: The influence of examiner citations. *The Review of Economics and Statistics*, 88(4): 774–779, 2006.

Juan Alcacer, Michelle Gittelman, and Bhaven Sampat. Applicant and examiner citations in us patents: An overview and analysis. *Research Policy*, 38(2):415–427, 2009.

Paul Almeida and Bruce Kogut. Localization of knowledge and the mobility of engineers in regional networks. *Management science*, 45(7):905–917, 1999.

Mathew Avery. Continuing abuse of the hatch-waxman act by pharmaceutical patent holders and the failure of the 2003 amendments. *Hastings LJ*, 60:171, 2008.

Pierre Azoulay, Joshua S Graff Zivin, and Bhaven N Sampat. The diffusion of scientific knowledge across time and space: Evidence from professional transitions for the superstars of medicine. Technical report, National Bureau of Economic Research, 2011.

Sharon Belenzon and Mark Schankerman. Spreading the word: Geography, policy, and knowledge spillovers. *Review of Economics and Statistics*, 95(3):884–903, 2013.

Antonin Bergeaud, Yoann Potiron, and Juste Raimbault. Classifying patents based on their semantic content. *PloS one*, 12(4):e0176310, 2017.

David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.

Nicholas Bloom, Mark Schankerman, and John Van Reenen. Identifying technology spillovers and product market rivalry. *Econometrica*, 81(4):1347–1393, 2013.

Stefano Breschi and Francesco Lissoni. Knowledge spillovers and local innovation systems: a critical survey. *Industrial and corporate change*, 10(4):975–1005, 2001.

Kevin Bryan, Yasin Ozcan, and Bhaven Sampat. A user's guide to in-text citations, 2018.

Kristy Buzard, Gerald A Carlino, Robert M Hunt, Jake K Carr, and Tony E Smith. Localized knowledge spillovers: Evidence from the agglomeration of american r&d labs and patent data. 2016.

Aaron Chatterji, Edward Glaeser, and William Kerr. Clusters of entrepreneurship and innovation. *Innovation Policy and the Economy*, 14(1):129–166, 2014.

Wesley M Cohen and Daniel A Levinthal. Absorptive capacity: A new perspective on learning and innovation. In *Strategic Learning in a Knowledge economy*, pages 39–67. Elsevier, 2000.

Christopher A Cotropia, Mark A Lemley, and Bhaven Sampat. Do applicant patent citations matter? *Research Policy*, 42(4):844–854, 2013.

Matthew Hoffman, Francis R Bach, and David M Blei. Online learning for latent dirichlet allocation. In *advances in neural information processing systems*, pages 856–864, 2010.

Adam B Jaffe. Technological opportunity and spillovers of r&d: Evidence from firms' patents, profits, and market value. *American Economic Review*, 76(5):984–1001, 1986.

Adam B Jaffe. Real effects of academic research. *The American economic review*, pages 957–970, 1989.

Adam B Jaffe and Gaétan De Rassenfosse. Patent citation data in social science research: Overview and best practices. *Journal of the Association for Information Science and Technology*, 68(6):1360–1374, 2017.

Adam B Jaffe, Manuel Trajtenberg, and Rebecca Henderson. Geographic localization of knowledge spillovers as evidenced by patent citations. *the Quarterly journal of Economics*, 108(3):577–598, 1993.

Adam B Jaffe, Manuel Trajtenberg, and Michael S Fogarty. The meaning of patent citations: Report on the nber/case-western reserve survey of patentees. Technical report, National bureau of economic research, 2000.

Sarah Kaplan and Keyvan Vakili. The double-edged sword of recombination in breakthrough innovation. *Strategic Management Journal*, 36(10):1435–1457, 2015.

Ryan Lampe. Strategic citation. *Review of Economics and Statistics*, 94(1):320–333, 2012.

Quoc Le and Tomas Mikolov. Distributed representations of sentences and documents. In *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, pages 1188–1196, 2014.

Josh Lerner and Amit Seru. The use and misuse of patent data: Issues for corporate finance and beyond. *Booth/Harvard Business School Working Paper*, 2015.

Chu-Cheng Lin, Waleed Ammar, Chris Dyer, and Lori Levin. Unsupervised pos induction with word embeddings. *arXiv preprint arXiv:1503.06760*, 2015.

Jeffrey Lin et al. The paper trail of knowledge transfers. *Federal Reserve Bank of Philadelphia Business Review. Second Quarter*, 2014.

Alan C Marco, Asrat Tesfayesus, and Andrew A Toole. Patent litigation data from us district court electronic records (1963-2015). 2017.

Aditi Mehta, Jessica K Martin, and Charlotte B Kahn. City of ideas: Reinventing boston's innovation economy: The boston indicators report 2012. 2012.

Rada Mihalcea, Courtney Corley, Carlo Strapparava, et al. Corpus-based and knowledge-based measures of text semantic similarity. In *AAAI*, volume 6, pages 775–780, 2006.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013a.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119, 2013b.

Petra Moser, Joerg Ohmstedt, and Paul W Rhode. Patent citations - an analysis of quality differences and citing practices in hybrid corn. *Management Science*, 64(4):1926–1940, 2017.

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

Mikko Packalen and Jay Battacharya. Cities and ideas, 2015.

Radim Řehůřek and Petr Sojka. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pages 45–50, Valletta, Malta, May 2010. ELRA. http://is.muni.cz/publication/884893/en.

Michael Roach and Wesley M Cohen. Lens or prism? patent citations as a measure of knowledge flows from public research. *Management Science*, 59(2):504–525, 2013.

Nathan Rosenberg. *Inside the black box: technology and economics*. Cambridge University Press, 1982.

Jung Won Sonn and Michael Storper. The increasing importance of geographical proximity in knowledge production: an analysis of us patent citations, 1975–1997. *Environment and Planning A*, 40(5):1020–1039, 2008.

Peter Thompson. Patent citations and the geography of knowledge spillovers: evidence from inventor-and examiner-added citations. *The Review of Economics and Statistics*, 88(2):383–388, 2006.

Peter Thompson and Melanie Fox-Kean. Patent citations and the geography of knowledge spillovers: A reassessment. *American Economic Review*, pages 450–460, 2005.

Stefan Wagner, Karin Hoisl, and Grid Thoma. Overcoming localization of knowledge - the role of professional service firms. *Strategic management journal*, 35(11):1671–1688, 2014.

Jeffrey M Wooldridge. *Introductory econometrics: A modern approach*. Nelson Education, 2015.