PRELIMINARY DRAFT

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

Fiona Sijie Feng (NYU Stern)

September 24, 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 0-0.08 standard deviations more similar than patents from different cities. Including further technology controls reduces estimates to -0.02 to 0.04. This suggests that local differentiation in technology space may play a minor role in determining innovation. 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. These findings indicate that geographic localization of knowledge spillovers may not be a strong driver of local innovation and agglomeration, as suggested by standard citations-based analyses.

1. Introduction

Knowledge spillovers are considered a key driving force of endogenous growth. 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. Consensus is that knowledge spillovers are geographically localized, meaning that local inventors and firms benefit more from the positive externalities generated by the R&D of other local firms, compared to inventors and firms located in different cities.

This paper uses methodology from unsupervised machine learning to analyze the ideas in patent abstracts, which summarize the invention. I make three key contributions: first, I derive numeric vector space representations of innovative ideas using the Document Vectors (DocVec) algorithm. Second, I extend the original research of JTH as a benchmark measure. Third, instead of using citations, I proxy for knowledge spillovers by analyzing the similarity of patents within technology fields.

I find that the evidence for localization is mixed. The citations-based benchmark finds that (i) localization estimates are large and significant: local patents receiving 0.24-0.30 SD more local citations compared to non-local patents; (ii) localization is increasing over the time frame 1975-2005. Similarity measures finds that (i) localization estimates are much smaller: local patents are -0.02 to 0.08 SD "more" similar to other local patents, after controlling for technological proximity; (ii) localization has declined 1995-2015.

After controlling for MSA-specific technological proximity, I find that local patents are on average 0.02 SD less similar to other local patents. This effect is particularly pronounced for patent pairs from highly similar sub-fields, that is, patents from local rivals. This implies that *local differentiation* may also play a role in the innovative process, where inventors distance themselves from other local inventors so as to broaden the scope of their patents. Such differentiation in innovation may also be related to product market rivalry (Bloom et al. (2013)).

The differences in the findings using citations and similarity knowledge spillover measures may be due to the influence of external knowledge sources, as inventors do not only learn from other inventors and their patents. This may include workplace training, technical journals, websites, magazines, textbooks and other publications. Focusing exclusively on citations will miss knowledge flows from outside sources, while patent text will capture a wider range of knowledge flows relevant for the innovation process. 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% of backward citations in common.

There are also well-documented concerns that citations themselves may not be good proxies for the relevant patent-based knowledge flows. Cotropia et al. (2013) and Alcacer et al. (2009) detail the extent to which patent examiners add their own citations they consider to be relevant prior art for the applicant's invention. Lampe (2012) and Roach and Cohen (2013) find evidence that citations are made strategically in ways that do not reveal knowledge flows, at least to other firms' privately held patents. I find evidence that support the previous literature. First, I examine the inventors' self-citation rates before and after they change firms. I find that inventors consistently cite their own prior inventions less after shifting firms, particular for relatively highly similar previous patents. This may be explained by the strategic incentive for firms to protect the scope of their inventions and thus leave out related citations of which they may be "reasonably ignorant". Second, I examine the change in patents cited by the inventor after they change firms. I find that inventors work on fairly similar inventions after changing firms (average similarity from 0.29 to 0.25), their list of cited patents changes drastically (from 16 citations in common to 2). This further suggests that citations may be determined by the firm's (or firm's lawyers') prior knowledge of other patents, rather than the indicating the relevant patents required for the innovation.

These findings imply that current methods of measuring knowledge spillovers may overestimate localization by focusing only on citation networks. While citation networks may be localized, 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.

Literature Knowledge spillovers are widely believed to be one of the key contributors to the phenomena of agglomeration. Krugman (1991) and Marshall and Marshall (1920) cite knowledge spillovers as a source of positive externality, but not necessarily confined to cities themselves. The localized nature of knowledge spillovers (i.e. the spread of knowledge is bounded within a geographic region) has been argued by Jacobs (1969), Manski (2000) and Glaeser et al. (1992), who observe that "intellectual breakthroughs must cross hallways and streets more easily than oceans and continents." Feldman (1994) motivates this by suggesting that firms in the same location reduce the uncertainty of innovation by sharing knowledge.

The empirical knowledge spillovers literature began with Jaffe et al. (1993). JTH examines whether or not inventors are more likely to cite patents closer to them, compared with a control patent. *To* separate localization effects from the existing distribution of patenting activity across locations, JTH select a control patent from the same technology distribution using PTO classes. They argue that since patents from the same PTO class come from the same technology distribution, then if there were no localization effects, there would not be significant differences in their citations' distribution across locations. Their study finds highly significant localization effects using the control/treatment methodology.

There are three main concerns with their methodology: first, whether or not their control selection method (using PTO primary class) is appropriate. Thompson and Fox-Kean (2005) use PTO primary subclass and find less significant localization effects; Murata et al. (2014) draw counterfactuals from a distribution, and alongside Buzard et al. (2016), use a continuous distance measure. Both papers find substantial evidence supporting localization. Instead of drawing counterfactual patents from PTO classes, another strand of literature uses the geographic mobility of inventors and authors to identify local spillover effects. Almeida and Kogut (1999), Agrawal et al. (2006), and Azoulay et al. (2011) all find that citations located in the inventor's or author's previous location is much higher than the control's citation rate. (These studies do still use the control/treatment method, just with stayers/movers) However, this does not address the second concern: that location choice cannot be taken as exogenous. As highlighted by Alcacer and Chung (2007), firms are strategic in their choice of location and may choose location based on where local spillover effects may be strong. Greenstone et al. (2010) directly address this by using "Million Dollar Plants" data to compare the agglomeration spillover effects in "winner" counties that were chosen as well as "runner-up" counties who narrowly lost, who serve as counterfactuals. They find evidence in favor of local TFP spillovers, though not directly addressing shared knowledge. Spatial econometrics have also been applied in addressing geographic localization. Following Bloom et al. (2013), Lychagin et al. (2016) find significant positive localization effects of technological proximity using vectorizations of PTO classes.

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. Section 5 presents the estimation models and results for localization regressions. Section 6 examines the influence of external sources of knowledge through new technology terms appearing in the patent corpus.

2. Previous Literature and Data Construction

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.

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.

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. 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 terms of finding highly similar documents, I focus my discussion on Doc2Vecs.

Unsupervised Machine Learning Methods: Document Vectors

Document Vectors (DocVecs) was introduced by Le and Mikolov (2014), and is a straightforward extension of the word2vec model of Mikolov et al. (2013b,a). Word2vec was developed in order to address issues with the bag-of-words procedure, primarily the loss of information provided by word ordering, and its inability to identify similar terms. To address this, word2vec uses the "context" around each term in the document in order to represent a term in vector space. For example, for the sentence "Provides for unattended file transfers", the word "unattended" has the context ["Provides", "for", "file", "transfers"]. Every word and

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

document is assigned a vector of dimension $N = 100$.[2] The resulting vector places words that arise in similar contexts close to each other. Since documents are treated more or less just like words, the algorithm places documents that contain similar terms close to each other in the vector space. See A.1.1 for more details on the algorithm; figure A.2 illustrates diagramatically the inputs and outputs of the algorithm. The main difference between Document Vectors and other representations is that the columns of the vectors do not represent anything meaningful; its purpose is solely to find a way to place documents in a vector space such that distances across the vectors represent similarity across patents as detected by the algorithm.

Unsupervised Machine Learning Methods: Topic Models

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 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.

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

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 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.

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 $[-1, 1]$.

Technology Controls: Cross Field Similarity

Since each patent is assigned technology field labels, the vector representation of a technology field f at t can be the *median*[4] of all patents within the field granted from year $t-5$ to t:

$$FV_{f,t} = median(PV_i | i \in f)_{t-5,t} \tag{2.2}$$

Further, I can treat the fields at each location as a "sub"-field, and define a location field vector at some MSA l as:

$$LFV_{f,l,t} = median(PV_i | i \in f, i \in l)_{t-5,t} \tag{2.3}$$

For LDA, this represents the distribution of all patents in the field across each of the 60 topics. For Document Vectors, this represents the centroid of the set of all patents within the field. For any patent i

[3] Other measures, such as Hellinger distance, were also used but found to be very highly correlated with cosine similarity.
[4] The median is chosen as a better representation of the field as it is less sensitive to outlier values.

in field f_i, cross-field similarity and cross-location-field similarity to patent j is analogously:

$$sim(f_i, f_j) = \frac{FV_{f_i} \cdot FV_{f_j}}{\|FV_{f_i}\| \, \|FV_{f_j}\|} \tag{2.4}$$

Intuitively, this is 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 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.[5]

What does patent similarity measure?

A more nuanced challenge is in the assessment of which measure is a better proxy for knowledge spillovers. 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 all of 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 similar knowledge influences, although such influences may not be directly linked. 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

[5] 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.

Year Group	NAICS Match	Year Match	Primclass Match	Inventor Match
1975-85	0.125	0.125	0.190	0.310
1985-95	0.122	0.122	0.186	0.334
1995-05	0.127	0.127	0.196	0.323
2005-15	0.139	0.140	0.200	0.317

Table 2.1: Average DocVecs Similarity Within Groups

the exchange of knowledge flows. Since we have not had an alternative to measuring spillovers until this point, the theoretical case for which measure is better suited for the task remains to be determined and is beyond the scope of this paper.

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. 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.

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. 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).[6] 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 noise from bins with few observations[7], the lowest bin includes all values below, and the highest bin includes all values above. Over

[6] To avoid truncation bias, only citations granted within the previous 10 years of the new patent were counted.
[7] Below the 1st percentile and above the 99th

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 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.

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. 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

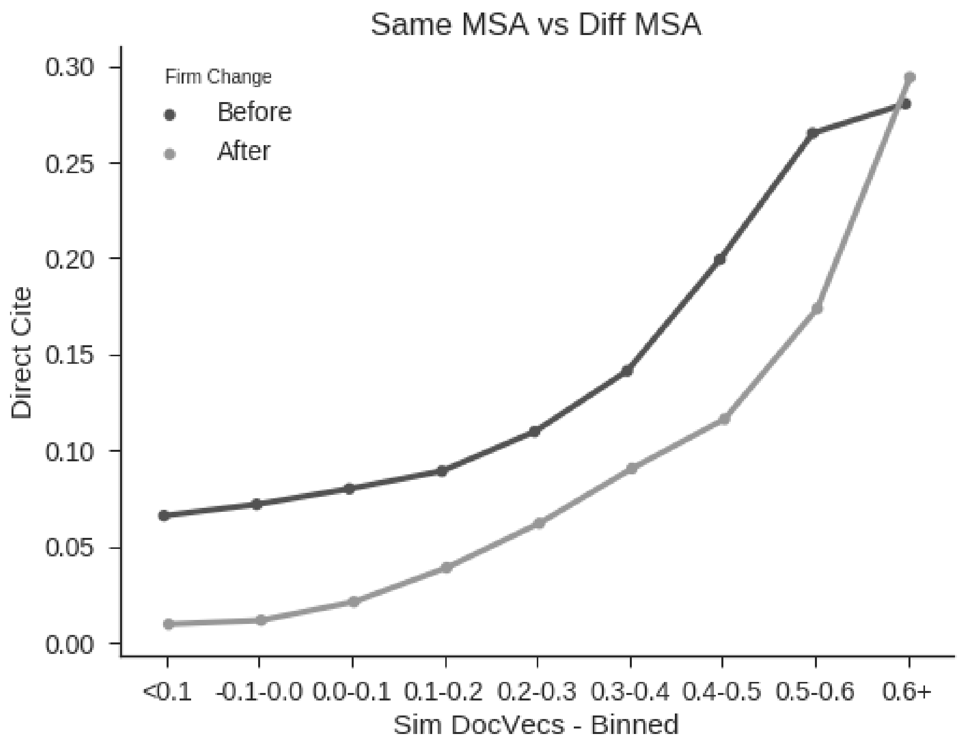


Figure 4.1: Rates of direct citation by DocVecs similarity. See C.1 for table of results.

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

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.

sample, I then examine how many citations are shared before and after the inventor changes firms. Citations made to other patents assigned to the same firm are excluded prior to the analysis.[8] I find that changing firms 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 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. 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.

[8]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.

	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

5. Measuring Local Knowledge Spillovers

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 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. Here, a citation is a patent that cites the target.

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]

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,

[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.

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

To isolate the effect of localization, the above exercise can be represented as a regression model:

$$pct\, cites\, in\, MSA_T = \beta_0 + \beta_1 I(MSA_i = MSA_T) + Year\, FE + \epsilon \tag{5.1}$$

Primcary class fixed effects are also added:

$$pct\, cites\, in\, MSA_T = \beta_0 + \beta_1 I(MSA_i = MSA_T) + Year\, FE + \epsilon \tag{5.2}$$

Where $i \in \{T = target, C = control\}$. Here, $I(MSA_T = MSA_T) = 1$, $I(MSA_C = MSA_T) = 0$. Then, as in **??**, the estimated measure of localization is $\hat{\beta}_1$. In this case, $\hat{\beta}_1$ represents the (approximate) increase in the percentage of citations in a particular MSA for a patent in the same MSA vs from different MSAs.

5.2. Changing the spillover measure to similarity

With similarities, we do not necessarily require a control patent as a point of comparison. 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. 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. Instead of using a control patent as baseline, I instead take the cross patent similarity of a target patent with citations from its own MSA, to citations in different MSAs. This approach has a straightforward interpretation: localization is present if a patent is more similar to its forward citations from the same MSA compared to those from different MSAs. This would mean that of the subsequent patents directly influenced by the target, on average, patents from the same location as the target share more in common.

I find that target patents *are* more similar to citations matching its own MSA, compared to citations from other MSAs. 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; for LDAVecs, same MSA citations are about 10% more similar than citations from different MSAs.

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

Grant Year	$\overline{sim_{LDA}(T,j \mid j\,in\,MSA_T)}$	$\overline{sim_{LDA}(T,j \mid j\,not\,in\,MSA_T)}$	Diff	t-stat	p-value
1975-85	0.58	0.53	0.05	17.81	0.0
1985-95	0.57	0.51	0.06	27.66	0.0
1995-05	0.54	0.49	0.05	33.25	0.0
2005-15	0.53	0.48	0.05	20.79	0.0

Table 5.2: JTH Replication Results with Similarity Measures.

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. In regression form, I estimate:

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

$$\overline{sim(T,j)} = \beta_0 + \beta_1 I(MSA_j = MSA_T) + Year\,FE + Primclass\,FE + \epsilon \tag{5.4}$$

$\hat{\beta}_1$ measures the proportional increase in the average similarity of a patent to citations in the same MSA compared to the similarity to citations in different MSAs, subject to controls.

5.3. Using random patent-pair similarity to measure spillovers and localization

As outlined in section 3, citations may omit many possible transmissions of knowledge. 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. [11]Including random pairs captures potential shared knowledge for any pair of patents and not only those linked by citation relationships, which represent only a small fraction of possible patent pairs. The hypothesis becomes: *knowledge spillovers are localized if patent pairs within a field are more similar if they are from the same MSA compared to different MSAs*. Focusing on patent pairs also allows me to bypass issues related to composition with more aggregated measures, and make these results more directly comparable with the JTH-based exercises above. 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 (cite TODO) is to find the number of common cited patents that appear as backward citations for both i and j.[12] Thus if $i$cites patents A, B, C, and j cites patents B, C, D, then the number of common cited patents is two. In this case, localization is significant if the number of common cited patents is higher between patent pairs from the same location.

Accounting for technological proximity and existing relationships between patents

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, even within JTH, who state: "... if a large fraction

[11]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.

[12]After self-citations are removed.

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." I accommodate this possibility in the second case: to better control for the existing technological proximity I also use cross-field similarity and cross-location-field similarity. Knowledge overlap may occur just through being in the same or similar technological fields; that the 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.

On the other hand, a well-known problem in the literature is that identification may be confounded as firms from similar technology fields also collocate in order to take advantage of potential spillovers, alongside other agglomeration benefits. Thus, controlling for the similarity across technology fields may be excessive. However, if we use the technology similarity control as a proxy for these other agglomeration effects, then any remaining knowledge localization effects is more reliably attributable to knowledge spillovers. This *strict* view may give us some sense of the lower bound for the size of localization.

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 sub-classes, 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 field vectors as described in equation (2.3). This treats each field 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.[13] These measures should also be a more "precise" approximation of the underlying patents, which further 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.

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

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

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." Additionally, if the patents are from different companies that share a lawyer, we may expect the lawyer to word inventions in a similar "style".

For the sampling procedure,[14] I sample a set of target patents, and pair these patents with both patents from the same field and MSA, and patents just within the same field. This is 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 that 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.[15]

Regression models for general localized knowledge spillovers

For general controls, I add year fixed effects and technology field fixed effects. The complete sample size for all years are approximately 1.5 million pairs for each sample. To allow for the effects to change over time, I group each sample into 10 year cross sections.

For both each measure, I begin by estimating the regression model without technology controls, where i, j are patent pairs within the same technology field:

$$sim(i, j) = \beta_0 + \beta_1 I(MSA_i = MSA_j) + Year\,FE + \epsilon \tag{5.5}$$

Then I add technology field fixed effects:

$$sim(i, j) = \beta_0 + \beta_1 I(MSA_i = MSA_j) + Year\,FE + Field\,FE + \epsilon \tag{5.6}$$

As a point of direct comparison of "similarity" across two patents using citations, I also measure the number of common cited patents between i, j. Self citations made to other patents assigned to the same firm are removed respectively. The analogous regression to equation (5.6) is then:

$$num\,common\,cited_{i,j} = \beta_0 + \beta_1 I(MSA_i = MSA_j) + Year\,FE + Field\,FE + \epsilon \tag{5.7}$$

These equations are directly comparable to JTH regressions equation (5.1) andequation (5.3). The results from these regressions are reported in table 5.4.

[14]See **??** for full outline

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

5.3.1. Regression models for strict localized knowledge spillovers

For the within NAICS sample,[16] pre-existing similarity across primary classes can be accounted for:

$$sim(i,j) = \beta_0 + \beta_1 I(MSA_i = MSA_j) + \beta_2 sim(pc_i, pc_j) + Year\,FE + Primclass\,FE + \epsilon \tag{5.8}$$

If patent i was granted in 2007 and had primary class *398: Optical communications*, the vector pc_i would represent the median of all patents with primary class *398* granted between 2001-2006.

I also allow for the localization effect to affect the slope coefficient β_3, which would imply that the similarity between patents in the same location *grows more* with the similarity in their primary classes:

$$\begin{aligned} sim(i,j) &= \beta_0 + \beta_1 I(MSA_i = MSA_j) + \beta_2 sim(pc_i, pc_j) \\ &\quad + \beta_3 I(MSA_i = MSA_j) * sim(pc_i, pc_j) + Primclass\,FE \\ &\quad + Year\,FE + \epsilon \end{aligned} \tag{5.9}$$

MSA-primary class subfields may also be used:

$$\begin{aligned} sim(i,j) = \beta_0 + \beta_1 I(MSA_i = MSA_j) + \beta_2 sim(pc_{i,MSA_i}, pc_{j,MSA_j}) \\ + Year\,FE + Primclass\,FE + \epsilon \end{aligned} \tag{5.10}$$

$$\begin{aligned} sim(i,j) &= \beta_0 + \beta_1 I(MSA_i = MSA_j) + \beta_2 sim(pc_i, pc_j) \\ &\quad + \beta_3 I(MSA_i = MSA_j) * sim(pc_{i,MSA_i}, pc_{j,MSA_j}) + Primclass\,FE \\ &\quad + Year\,FE + \epsilon \end{aligned} \tag{5.11}$$

If patent i was granted in Los Angeles, pc_{i,MSA_i} represents the median of all patents granted in primary

[16] These regressions can not be applied to the within-primary class sample, as similarity across primary class would be degenerate. Similarity across MSA-primary class pairs is also degenerate when $I(MSA_i = MSA_j) = True$, so it is not possible to separately identify β_1.

class 398 in the city of Los Angeles, from 2001-2006. The interpretation here would be that if $\hat{\beta}_1, \hat{\beta}_2$ or $\hat{\beta}_3$ are positive and significant, then innovations from the same location share more in common compared to other patent pairs from technology fields that are just as similar. For example, suppose $pc_i = 398$ and $MSA_i = Los\,Angeles$; $pc_j = 396$ and $MSA_j = Austin$; $pc_k = 396$ and $MSA_k = Los\,Angeles$. If the similarity of past patents from primary class 398 in Los Angeles to primary class 396 from both Austin and Los Angeles, and yet the patent pairs i, k from Los Angeles are more similar on average to the i, j Los Angeles-Austin patent pairs, then localization is significant.

5.4. Results: Estimating Localization From Regression

Localized knowledge spillovers: Baseline Comparison

To increases the comparability of localization from each method, I normalize all measures of spillovers. Then the estimate of $\hat{\beta}_1$, represents the standard deviation increase in knowledge spillovers when patents are local. For the baseline comparison, I include year and primary class fixed effects in each set of regression models.[17]The main finding from table 5.4 is 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. [18] 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 E.3) 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 estimates using the number of common cited patents are also much smaller in magnitude compared to JTH's citation measure. However, the time trend matches: both measures find overall increasing localization effects for the observed time period. Localization estimates rise from 0.002 to 0.05 over the period 1975-2015 for the within-NAICS sample; from 0.01 to 0.11 for within-Primclass. This

[17]The same results with raw untransformed data is reported in table **??**, alongside results for LDAVecs (table E.4,table E.5). In the vast majority of cases between LDAVecs and DocVecs, the sign and significance of the the coefficient estimates align, while some magnitudes may differ.

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

agrees with DocVecs estimates that localization effects are higher for the Primclass sample. As patents are more similar to other local patents in narrower fields of technology. This indicates that local knowledge sharing may happen more in specialized areas of innovation, compared to knowledge sharing at the industry level.

$KS =$ Pct Cites in Target's MSA, JTH Sample			
	1975-85	1985-95	1995-05
$I(MSA\, Match)$	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
$I(MSA\, Match)$	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.

The low estimates for localization is not due to noise in the similarity measures. If we replace $I(MSA\, Match)$ with another indicator which we expect to have a large effect on the similarity, the estimate of the coefficient aligns with expectations are is much larger than the estimate for localization. Using the within-NAICS sample, If the patent pair shares an inventor $(I(Inv\, Match) = 1)$, then this increases similarity by 1.27 to 1.52 standard deviations. Similarly, if patents are from the same primary class $(I(Primclass\, Match) = 1)$, then similarity increases by 0.4-0.44 standard deviations. (See full results in E.2) Thus, the similarity measure is able to capture significant effects.

KS = Sim DocVecs, Within-NAICS				
	1975-85	1985-95	1995-05	2005-15
$I(MSA\,Match)$	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
$I(MSA\,Match)$	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 = Num Common Cited, Within-NAICS				
	1975-85	1985-95	1995-05	2005-15
$I(MSA\,Match)$	0.0019***	0.0074***	0.0155***	0.0514***
	(0.0003)	(0.0014)	(0.0018)	(0.0084)
N	194131	282112	443885	578056
Adjusted R^2	0.00	0.03	0.00	0.00
Year FE	True	True	True	True
PC FE	True	True	True	True

KS = Num Common Cited, Within-Primclass				
	1975-85	1985-95	1995-05	2005-15
$I(MSA\,Match)$	0.0113***	0.0300***	0.0513***	0.1109***
	(0.0012)	(0.0029)	(0.0045)	(0.0084)
N	171893	252886	407176	537878
Adjusted R^2	0.01	0.05	0.00	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.

Localized knowledge spillovers: technology proximity and other controls

I find that including prior similarity across primary classes as controls in the within NAICS sample reduces the estimate for localization further, so that local patent pairs are about 0.025-0.045 standard deviations higher than non-local pairs ($I(MSA\,Match) = False$). For the same time range but without the similarity controls, [19] the localization estimate is 0.05-0.06. The time trend indicates that localization is diminishing over the period 1985-2015. The interaction term is also found to be slightly positive and significant for 1985-2015, stable at around 0.026; this means that local patent similarity grows at a slightly higher rate with primary class similarity. The main results are presented in table 5.5; figure E.3 depicts the conditional means plot. The raw data (table E.3) differs slightly in that the coefficient on the intercept dummy (i.e. $\hat{\beta}_1$ in equation (5.8)) is slightly negative and significant, versus positive and significant in the normalized regressions.

Interestingly, including MSA-specific primary class controls estimates localization to be significant and slightly *negative*. Local patent pairs are found to be between 0.017-0.023 standard deviations less similar than non-local patent pairs. The interaction term is found to be insignificant 1985-2005, and slightly positive and significant in 2005-2015. One thing that contributes to the negative estimate of the dummy intercept term is that $sim(pc_i, pc_j)$ is *not* highly correlated with $I(MSA\,Match)$ at 0.04; while the pairwise correlation with $sim(pc_{i,MSA_i}, pc_{j,MSA_j})$ is 0.19.[20] In other words, patent pairs from similar MSA-technology subfields are much more likely to be collocated than those from similar primary classes. Since the average difference between the local and non-local samples are already low, this slight collinearity may reduce the estimated effect of $I(MSA\,Match)$ even further. Another view could be that the collocation of patents from similar MSA subfields already indicates some presence of spillovers; while this certainly indicates the presence of clustering, whether or not clustering is equatable to local spillovers is open to interpretation.

The conditional mean plots (figure E.3, figure E.4) also sheds light on this effect. The average similarity of non-local patent pairs is slightly higher than local pairs at *higher* levels of MSA-subfield similarity. What is interesting is that local similarity does not increase as subfield similarity rises from 2 standard deviations above the mean to 2.5; while non-local similarity does substantially (the caveat being there are far fewer non-local patent pairs in the 2.5 bin). This may indicate the presence of *local differentiation* in technology space. One aspect of local knowledge spillovers may be that knowing more about neighbouring rivals'

[19] 1975-1985 sample drops out due to lack of data on past primary class patents.
[20] The distributions of these two measures for local and non-local patent pairs are in figure E.1,figure E.2.

inventions may push inventors to distance themselves from one another. That is, inventors want their innovations to be novel from that of their peers, as patents are more valuable when they have a broader scope. These results show that learning from local rivals may not necessarily lead to greater technological affinity.

For patent pairs with subfield similarity above 2 s.d., approximately 19% are Pharmaceutical pairs and a further 14% are Other Chemical Product pairs. In the general sample, they comprise 3.3% and 4.9% respectively. In both industries, particularly pharmaceuticals, the patent to product progression is much more straight forward compared to most other industries. As patents in these industries behave more like products, the consideration of product market differentiation may induce greater distancing in technology space by local firms.

While the sign of the results are different for MSA-specific primary class controls, in terms of absolute magnitude they do not differ substantially. The explanatory power of the primary class similarity measure is also slightly greater, as is its estimated effect on patent pair similarity. One standard deviation increase in $sim(pc_i, pc_j)$ increases $sim(i, j)$ by approximately 0.24-0.26 standard deviations; while for $sim(pc_{i,MSA_i}, pc_{j,MSA_j})$ the increase is between 0.17-0.19. Thus, while primary class similarity appear to be a slightly better fit for the data, the exercise using MSA-primary class similarity reveals potentially deeper dynamics of the knowledge spillovers process.

It has been documented by Moser (2011); Arora (1997) that industries such as Chemicals rely more heavily on secrecy to preserve intellectual property which leads to higher rates of localization. On the other hand, industries dominated by software patents which are known to have more diffuse knowledge networks, thus lowering the localization effect. I also drop Machinery pairs as they also constitute a large proportion of the total sample. I find that dropping each of these industries from the sample do not affect the estimation results in table 5.4. Using assignee level fixed effects such a firm size also have no effect. Using data at the patent pair level and including technology fixed effects has made the results fairly robust.

The overall findings on the presence and size of localized knoweldge spillovers is mixed using similarity measures. While citations finds large and significantly positive localization effects, the findings with similarity is comparatively muted. The difference in these findings, as discussed in 3, could possibly be attributed to the influence of external sources of knowledge which are less localized than citations networks.

1985-95				
	(1)	(2)	(3)	(4)
$I(MSA\,Match)$	0.0399***	-0.0177***	0.0434***	-0.0169***
	(0.0042)	(0.0047)	(0.0043)	(0.0049)
$sim_{DV}(pc_i, pc_j)$	0.2475***		0.2408***	
	(0.0024)		(0.0026)	
$sim_{DV}(pc_{i,MSA_i}, pc_{j,MSA_j})$		0.1719***		0.1698***
		(0.0026)		(0.0033)
$I_{MSA} * sim_{DV}(pc_i, pc_j)$			0.0254***	
			(0.0041)	
$I_{MSA} * sim_{DV}(pc_{i,MSA_i}, pc_{j,MSA_j})$				0.0051
				(0.0046)
N	281018	227029	281018	227029
Adjusted R^2	0.11	0.09	0.11	0.09
Year FE	True	True	True	True
PC FE	True	True	True	True

1995-05				
	(1)	(2)	(3)	(4)
$I(MSA\,Match)$	0.0360***	-0.0226***	0.0371***	-0.0228***
	(0.0033)	(0.0036)	(0.0033)	(0.0036)
$sim_{DV}(pc_i, pc_j)$	0.2448***		0.2377***	
	(0.0019)		(0.0021)	
$sim_{DV}(pc_{i,MSA_i}, pc_{j,MSA_j})$		0.1919***		0.1905***
		(0.0020)		(0.0024)
$I_{MSA} * sim_{DV}(pc_i, pc_j)$			0.0261***	
			(0.0033)	
$I_{MSA} * sim_{DV}(pc_{i,MSA_i}, pc_{j,MSA_j})$				0.0035
				(0.0035)
N	437485	382972	437485	382972
Adjusted R^2	0.12	0.11	0.12	0.11
Year FE	True	True	True	True
PC FE	True	True	True	True

2005-15				
	(1)	(2)	(3)	(4)
$I(MSA\,Match)$	0.0327***	-0.0175***	0.0266***	-0.0228***
	(0.0030)	(0.0032)	(0.0030)	(0.0033)
$sim_{DV}(pc_i, pc_j)$	0.2553***		0.2481***	
	(0.0019)		(0.0021)	
$sim_{DV}(pc_{i,MSA_i}, pc_{j,MSA_j})$		0.1916***		0.1870***
		(0.0018)		(0.0021)
$I_{MSA} * sim_{DV}(pc_i, pc_j)$			0.0263***	
			(0.0034)	
$I_{MSA} * sim_{DV}(pc_{i,MSA_i}, pc_{j,MSA_j})$				0.0132***
				(0.0032)
N	569100	516531	569100	516531
Adjusted R^2	0.09	0.08	0.09	0.08
Year FE	True	True	True	True
PC FE	True	True	True	True

Table 5.5: Regression results for strict local spillovers with normed data for LDAVec similarity.

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 G.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 G.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 G.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.[21] 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% of backward citations made. figure G.4 displays a wide variety of similarity across the initial adenovirus patents, which had an average similarity of 0.26.

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

7. 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 inputs and influences to the innovative process and its knowledge outputs. This implies that studying the flows of knowledge using citations is not interchangeable with studying flows within knowledge content. This has profound impact for citations based research, which has relied on an assumption of equivalence between the two. I discussed reasons within the patent citations literature of what may be behind the discrepancies, largely due to the strategic manner in which citations are made in relation to the value of patents as intellectual property. However, I expect that this may be attenuated within the academic literature, where there are fewer incentives for strategic citations. More research should focus on when and why there are significant differences in the localization of knowledge inputs to knowledge outputs.

There are also implications for R&D policy: public investment in R&D has been advocated on the premise of the existence of large and significant local returns. These findings complicate the advantages that knowledge spillovers offers, in that they do not appear to benefit local firms and inventors significantly more so. Additionally, there is limited evidence that localization spurs further innovation within a city. If this is the goal, then R&D policy should be directed towards improving possibilities for novel and radical innovation, not necessarily the greater sharing of existing knowledge. Further research should be conducted to investigate the relationship between knowledge spillovers, innovation novelty, and innovation growth.

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.[22] 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.

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

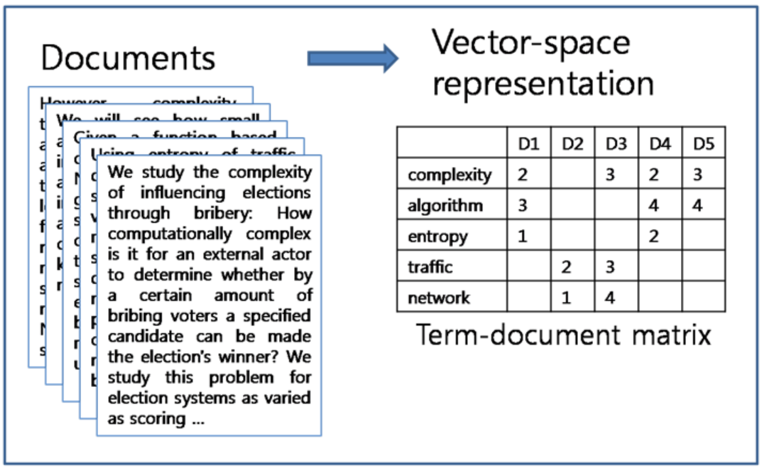


Figure A.1: Example of Document Term Matrix

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

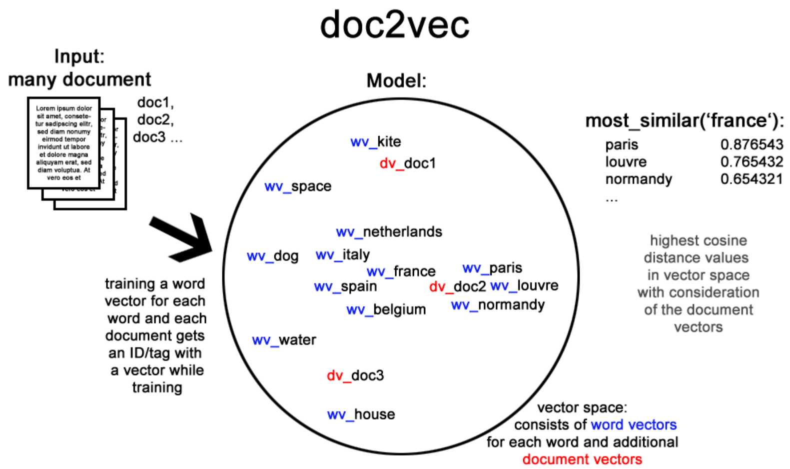


Figure A.2: Illustration of Document Vectors.

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, introduced by Blei et al. (2003), is a method of topic modeling which seeks to uncover the underlying structure of a group of documents (corpus) using just the observed text. It is a probabilistic model based on hierarchical Bayesian analysis. 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:

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

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.

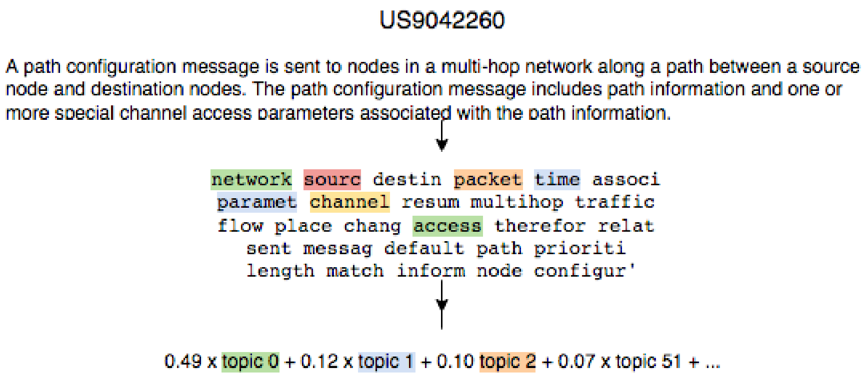


Figure A.3: Example of a patent converted into a distribution over topics.

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.

B. Citations and Patent Vector Similarity

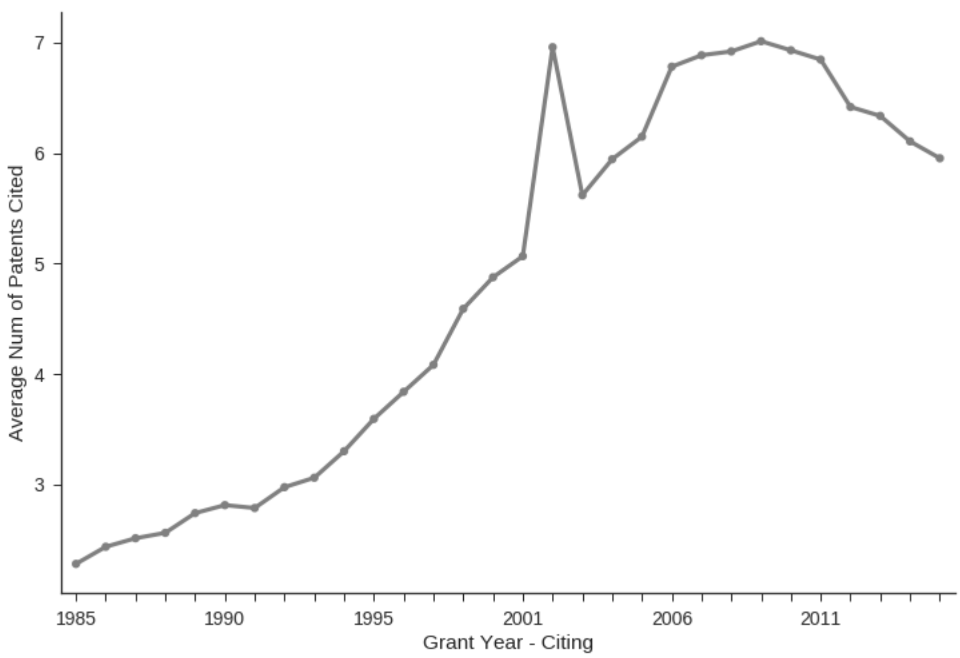


Figure B.1: Averate number of patents cited over time

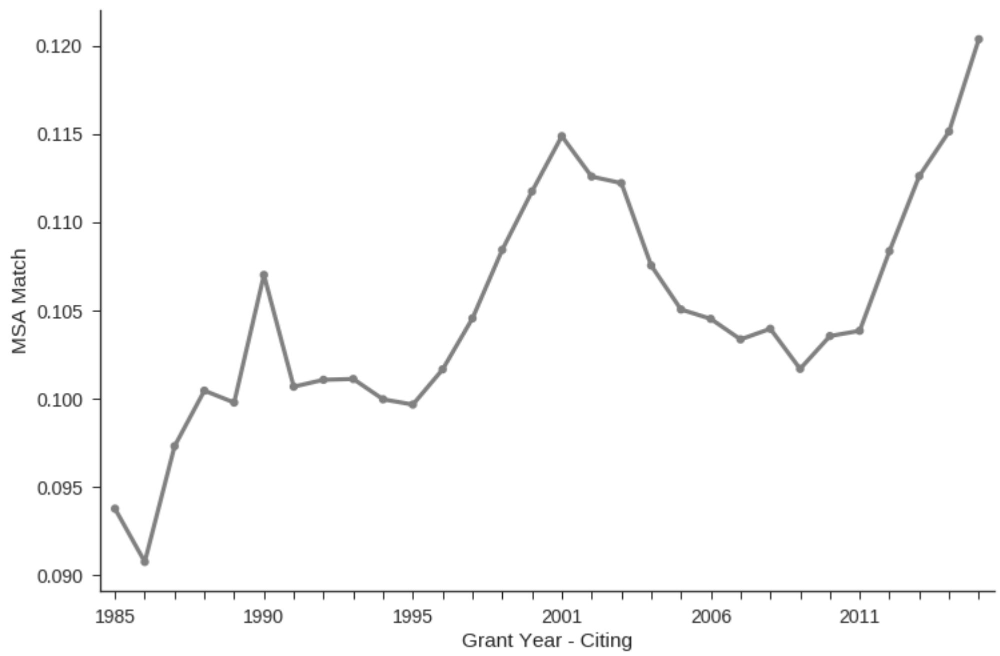


Figure B.2: Proportion of cited patents in the same MSA over time

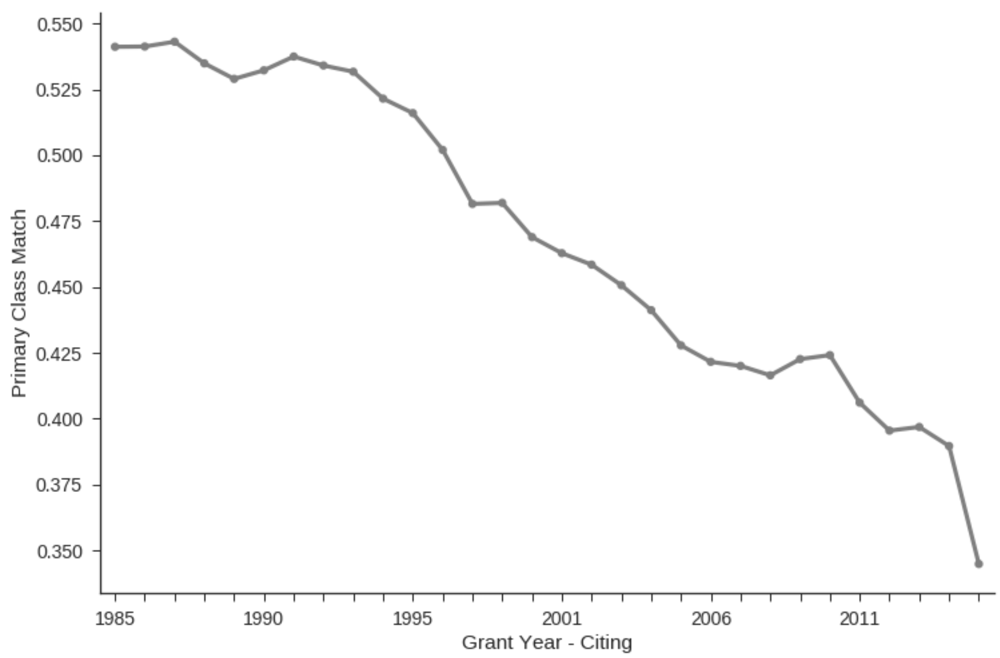


Figure B.3: Proportion of cited patents in the same primary class over time

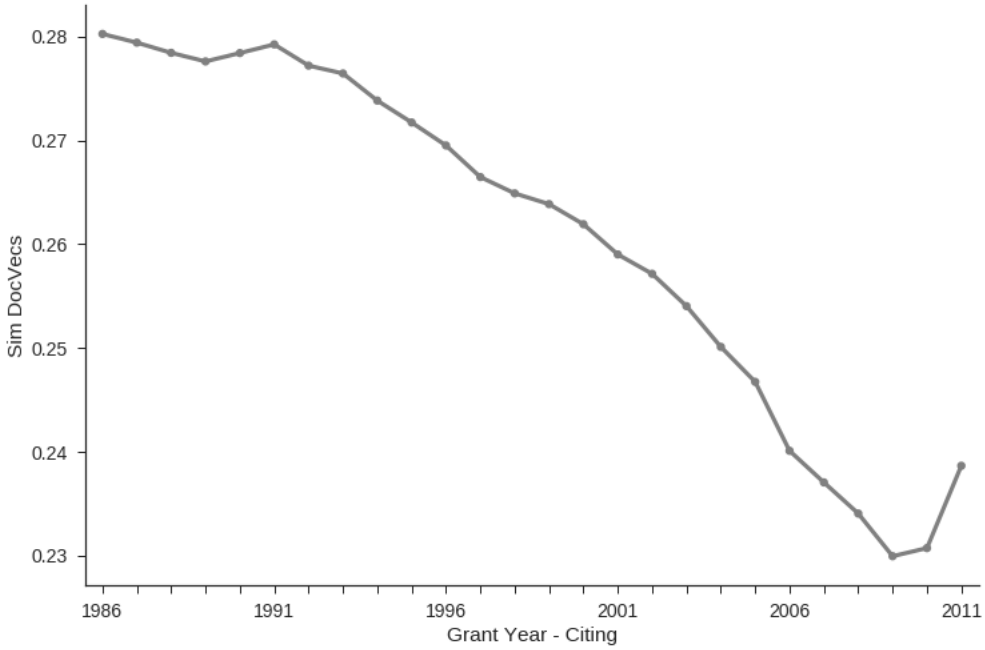


Figure B.4: Average DocVecs similarity to cited patents

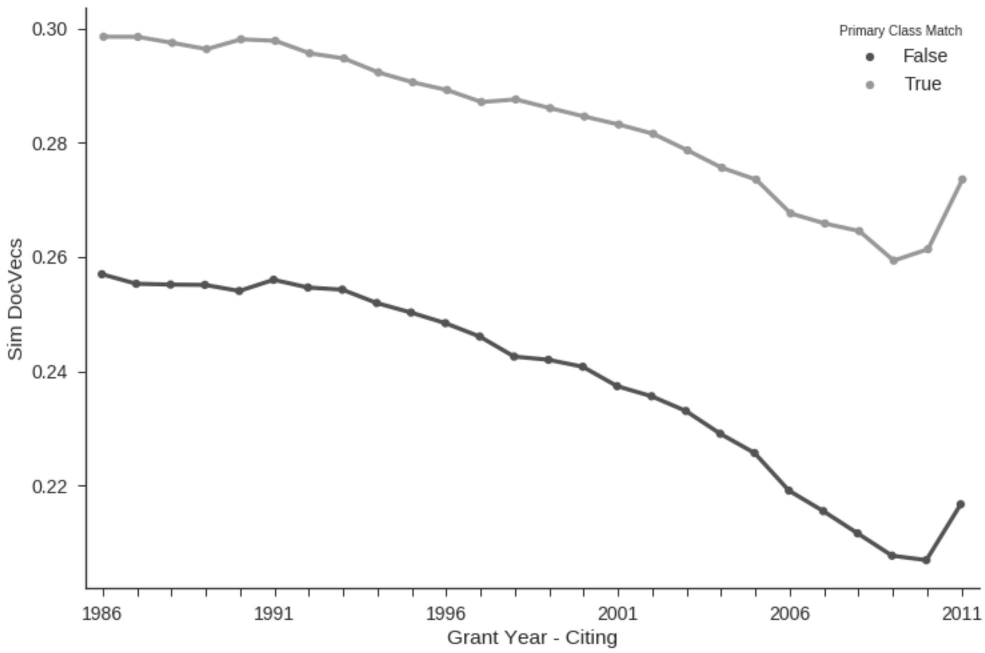


Figure B.5: Average DocVecs similarity to cited patents in the same primary class

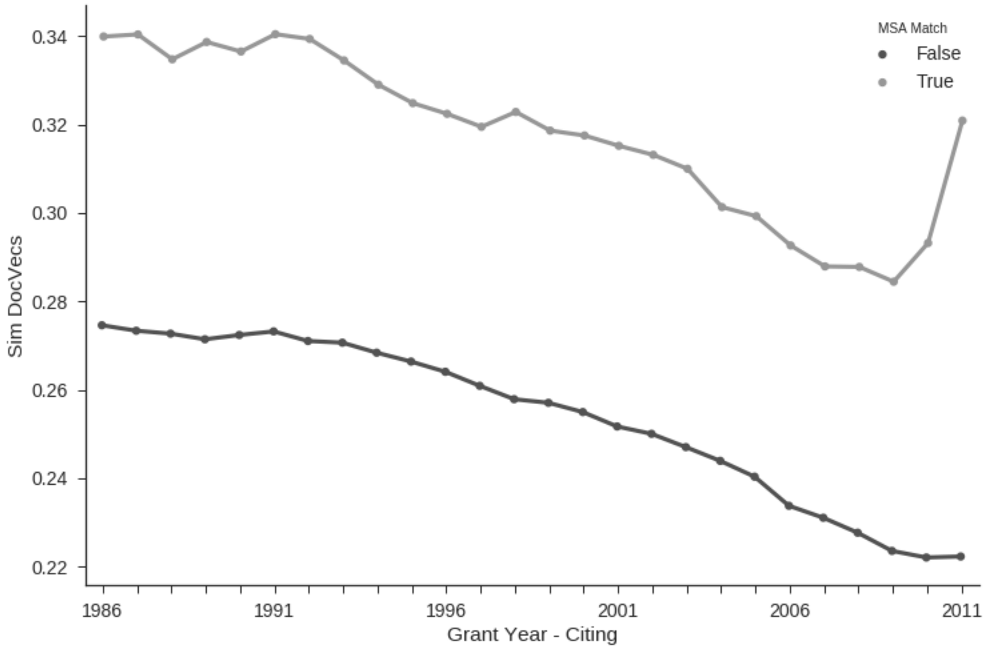


Figure B.6: Average DocVecs similarity to cited patents in the same MSA

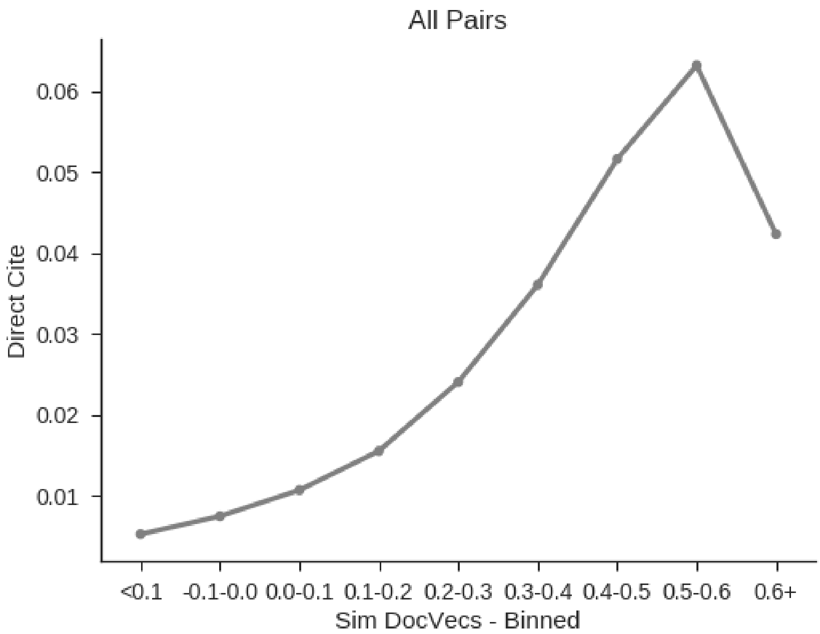


Figure B.7: Rate of direct citation conditional on level of DocVecs Similarity, All Pairs

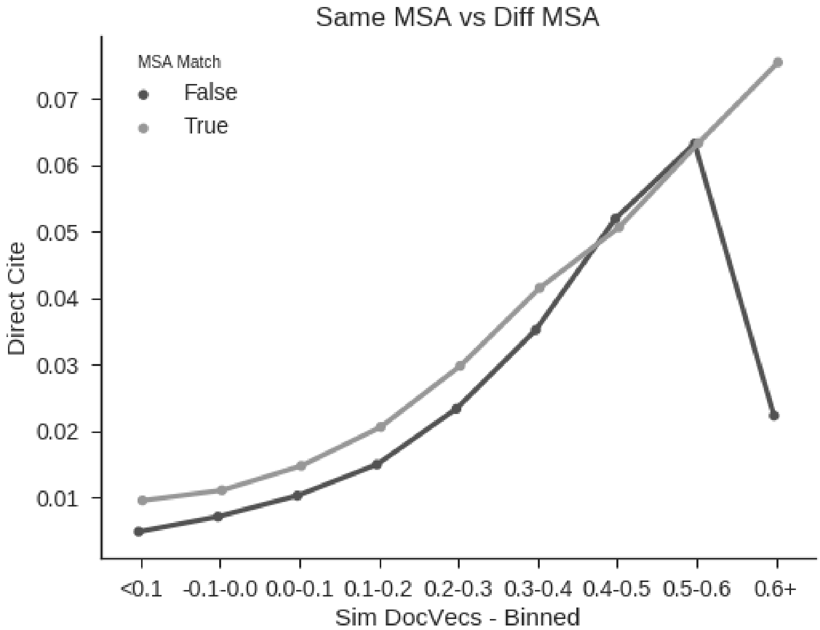


Figure B.8: Rate of direct citation conditional on level of DocVecs Similarity, Same MSA vs Different MSA

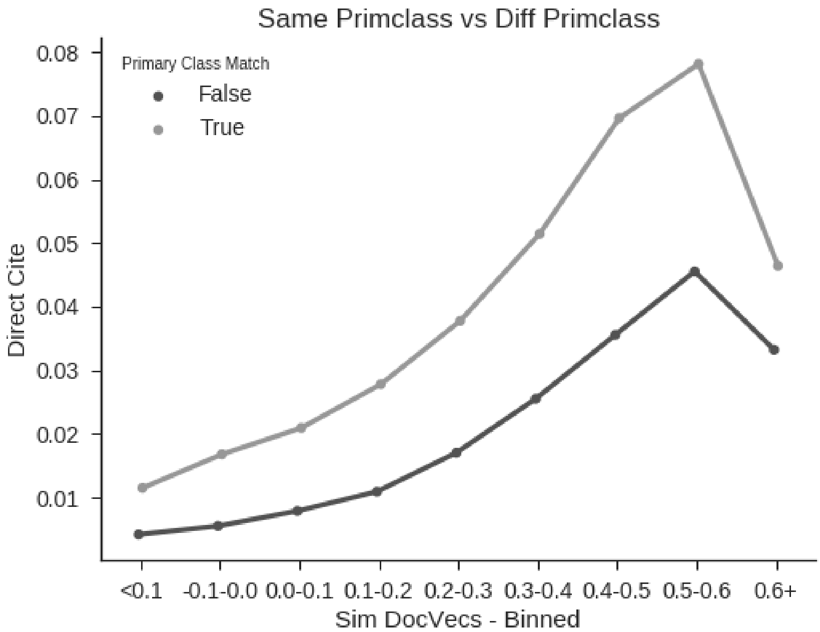


Figure B.9: Rate of direct citation conditional on level of DocVecs Similarity, Same Primary Class vs Different Primary Class

	<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

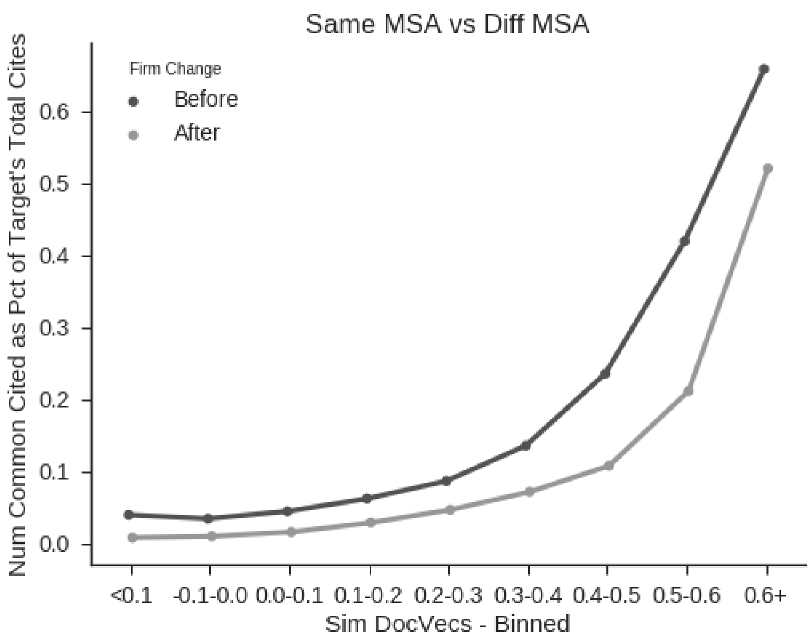


Figure C.1: Rate of direct citation conditional on level of DocVecs Similarity, Same Primary Class vs Different Primary Class

D. Summary of Data

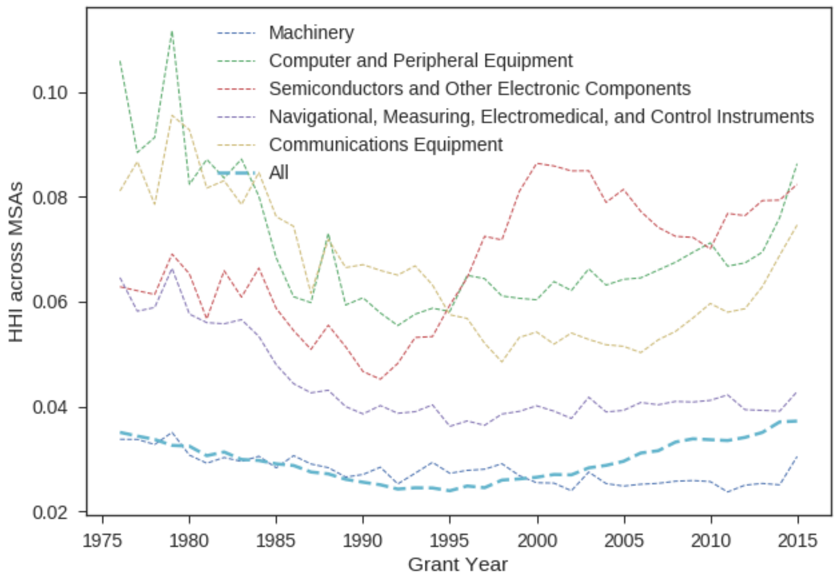


Figure D.1: HHI of concentration across cities for industry innovation

Figure D.2: HHI of concentration across cities for industry innovation

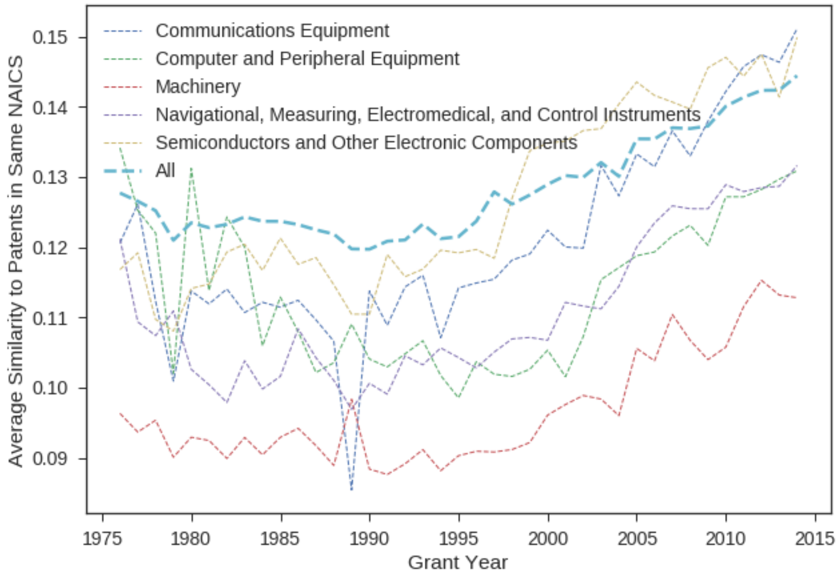


Figure D.3: DocVecs Similarity to other patents within the same industry over time

E. Regressions

Sample	Subset	y, KS	N	Mean	Std Dev	Min	25%	50%	75%	Max
JTH Extension	All	$pct\, cites\, in\, MSA_T$	507870	0.082	0.217	0.000	0.000	0.000	0.000	1.000
	$MSA\, Match = T$	$pct\, cites\, in\, MSA_T$	282079	0.112	0.250	0.000	0.000	0.000	0.042	1.000
	$MSA\, Match = F$	$pct\, cites\, in\, MSA_T$	225791	0.045	0.159	0.000	0.000	0.000	0.000	1.000
JTH Cite Sim	All	$\overline{sim_{LDA}(T,j)}$	283080	0.510	0.201	0.002	0.369	0.509	0.654	1.000
	$MSA\, Match = T$	$\overline{sim_{LDA}(T,j)}$	57897	0.550	0.222	0.002	0.392	0.555	0.721	1.000
	$MSA\, Match = F$	$\overline{sim_{LDA}(T,j)}$	225183	0.500	0.194	0.002	0.365	0.499	0.638	1.000
JTH Cite Sim	All	$\overline{sim_{DV}(T,j)}$	283080	0.289	0.124	-0.552	0.211	0.285	0.362	0.936
	$MSA\, Match = T$	$\overline{sim_{DV}(T,j)}$	57897	0.323	0.151	-0.552	0.224	0.314	0.409	0.936
	$MSA\, Match = F$	$\overline{sim_{DV}(T,j)}$	225183	0.280	0.114	-0.518	0.209	0.279	0.351	0.890
NAICS	All	$sim_{LDA}(i,j)$	1483214	0.242	0.222	0.001	0.061	0.174	0.368	1.000
	$MSA\, Match = T$	$sim_{LDA}(i,j)$	363370	0.255	0.228	0.001	0.067	0.188	0.389	1.000
	$MSA\, Match = F$	$sim_{LDA}(i,j)$	1119844	0.237	0.220	0.001	0.059	0.169	0.361	1.000
NAICS	All	$sim_{DV}(i,j)$	1481000	0.131	0.135	-0.416	0.039	0.123	0.213	0.679
	$MSA\, Match = T$	$sim_{DV}(i,j)$	362662	0.136	0.136	-0.416	0.043	0.128	0.219	0.679
	$MSA\, Match = F$	$sim_{DV}(i,j)$	1118338	0.129	0.134	-0.416	0.037	0.121	0.211	0.679
Primclass	All	$sim_{LDA}(i,j)$	1355828	0.377	0.245	0.001	0.172	0.347	0.559	1.000
	$MSA\, Match = T$	$sim_{LDA}(i,j)$	285175	0.389	0.249	0.001	0.180	0.360	0.577	1.000
	$MSA\, Match = F$	$sim_{LDA}(i,j)$	1070653	0.374	0.244	0.001	0.170	0.343	0.554	1.000
Primclass	All	$sim_{DV}(i,j)$	1354365	0.188	0.139	-0.371	0.094	0.183	0.275	0.748
	$MSA\, Match = T$	$sim_{DV}(i,j)$	284609	0.196	0.142	-0.367	0.100	0.190	0.284	0.748
	$MSA\, Match = F$	$sim_{DV}(i,j)$	1069756	0.186	0.138	-0.371	0.093	0.181	0.273	0.748

Table E.1: Summary statistics of knowledge spillover measures.

	1975-85	1985-95	1995-05	2005-15
$I(Primclass\, Match)$	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

	1975-85	1985-95	1995-05	2005-15
$I(Inv\, Match)$	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

Table E.2: Regression results replacing MSA Match with Inventor Match and Primclass Match.

1985-95				
	(1)	(2)	(3)	(4)
$I(MSA\,Match)$	0.0054***	-0.0024***	-0.0064***	-0.0037***
	(0.0006)	(0.0006)	(0.0019)	(0.0012)
$sim_{DV}(pc_i, pc_j)$	0.1748***		0.1700***	
	(0.0017)		(0.0018)	
$sim_{DV}(pc_{i,MSA_i}, pc_{j,MSA_j})$		0.0933***		0.0922***
		(0.0014)		(0.0018)
$I_{MSA} * sim_{DV}(pc_i, pc_j)$			0.0179***	
			(0.0029)	
$I_{MSA} * sim_{DV}(pc_{i,MSA_i}, pc_{j,MSA_j})$				0.0028
				(0.0025)
N	281018	227029	281018	227029
Adjusted R^2	0.11	0.09	0.11	0.09
Year FE	True	True	True	True
PC FE	True	True	True	True

1995-05				
	(1)	(2)	(3)	(4)
$I(MSA\,Match)$	0.0049***	-0.0031***	-0.0076***	-0.0040***
	(0.0004)	(0.0005)	(0.0015)	(0.0010)
$sim_{DV}(pc_i, pc_j)$	0.1728***		0.1678***	
	(0.0013)		(0.0015)	
$sim_{DV}(pc_{i,MSA_i}, pc_{j,MSA_j})$		0.1042***		0.1034***
		(0.0011)		(0.0013)
$I_{MSA} * sim_{DV}(pc_i, pc_j)$			0.0184***	
			(0.0023)	
$I_{MSA} * sim_{DV}(pc_{i,MSA_i}, pc_{j,MSA_j})$				0.0019
				(0.0019)
N	437485	382972	437485	382972
Adjusted R^2	0.12	0.11	0.12	0.11
Year FE	True	True	True	True
PC FE	True	True	True	True

2005-15				
	(1)	(2)	(3)	(4)
$I(MSA\,Match)$	0.0044***	-0.0024***	-0.0091***	-0.0066***
	(0.0004)	(0.0004)	(0.0017)	(0.0011)
$sim_{DV}(pc_i, pc_j)$	0.1803***		0.1751***	
	(0.0013)		(0.0015)	
$sim_{DV}(pc_{i,MSA_i}, pc_{j,MSA_j})$		0.1040***		0.1015***
		(0.0010)		(0.0012)
$I_{MSA} * sim_{DV}(pc_i, pc_j)$			0.0186***	
			(0.0024)	
$I_{MSA} * sim_{DV}(pc_{i,MSA_i}, pc_{j,MSA_j})$				0.0071***
				(0.0017)
N	569100	516531	569100	516531
Adjusted R^2	0.09	0.08	0.09	0.08
Year FE	True	True	True	True
PC FE	True	True	True	True

Table E.3: Regression results for strict local spillovers with raw data for DocVec similarity.

	1985-95			
	(1)	(2)	(3)	(4)
$I(MSA\,Match)$	0.0274***	-0.0415***	0.0280***	-0.0412***
	(0.0038)	(0.0042)	(0.0039)	(0.0042)
$sim_{LDA}(pc_i, pc_j)$	0.4123***		0.4094***	
	(0.0023)		(0.0026)	
$sim_{LDA}(pc_{i,MSA_i}, pc_{j,MSA_j})$		0.3994***		0.4221***
		(0.0026)		(0.0031)
$I_{MSA} * sim_{LDA}(pc_i, pc_j)$			0.0114***	
			(0.0044)	
$I_{MSA} * sim_{LDA}(pc_{i,MSA_i}, pc_{j,MSA_j})$				-0.0619***
				(0.0047)
N	281421	227396	281421	227396
Adjusted R^2	0.21	0.21	0.21	0.21
Year FE	True	True	True	True
PC FE	True	True	True	True
	1995-05			
	(1)	(2)	(3)	(4)
$I(MSA\,Match)$	0.0535***	-0.0202***	0.0533***	-0.0174***
	(0.0031)	(0.0033)	(0.0031)	(0.0032)
$sim_{LDA}(pc_i, pc_j)$	0.4349***		0.4286***	
	(0.0019)		(0.0021)	
$sim_{LDA}(pc_{i,MSA_i}, pc_{j,MSA_j})$		0.4219***		0.4366***
		(0.0019)		(0.0023)
$I_{MSA} * sim_{LDA}(pc_i, pc_j)$			0.0234***	
			(0.0034)	
$I_{MSA} * sim_{LDA}(pc_{i,MSA_i}, pc_{j,MSA_j})$				-0.0417***
				(0.0035)
N	438117	383569	438117	383569
Adjusted R^2	0.24	0.24	0.24	0.25
Year FE	True	True	True	True
PC FE	True	True	True	True
	2005-15			
	(1)	(2)	(3)	(4)
$I(MSA\,Match)$	0.0356***	-0.0279***	0.0343***	-0.0215***
	(0.0028)	(0.0029)	(0.0027)	(0.0028)
$sim_{LDA}(pc_i, pc_j)$	0.4299***		0.4274***	
	(0.0016)		(0.0018)	
$sim_{LDA}(pc_{i,MSA_i}, pc_{j,MSA_j})$		0.4096***		0.4197***
		(0.0016)		(0.0019)
$I_{MSA} * sim_{LDA}(pc_i, pc_j)$			0.0090***	
			(0.0028)	
$I_{MSA} * sim_{LDA}(pc_{i,MSA_i}, pc_{j,MSA_j})$				-0.0304***
				(0.0028)
N	569947	517333	569947	517333
Adjusted R^2	0.26	0.27	0.26	0.27
Year FE	True	True	True	True
PC FE	True	True	True	True

Table E.4: Regression results for strict local spillovers with normed data for LDAVec similarity.

1985-95				
	(1)	(2)	(3)	(4)
$I(MSA\,Match)$	0.0061***	-0.0092***	0.0038***	0.0044***
	(0.0008)	(0.0009)	(0.0010)	(0.0012)
$sim_{LDA}(pc_i, pc_j)$	0.2331***		0.2315***	
	(0.0013)		(0.0015)	
$sim_{LDA}(pc_{i,MSA_i}, pc_{j,MSA_j})$		0.2509***		0.2652***
		(0.0016)		(0.0020)
$I_{MSA} * sim_{LDA}(pc_i, pc_j)$			0.0064***	
			(0.0025)	
$I_{MSA} * sim_{LDA}(pc_{i,MSA_i}, pc_{j,MSA_j})$				-0.0389***
				(0.0030)
N	281421	227396	281421	227396
Adjusted R^2	0.21	0.21	0.21	0.21
Year FE	True	True	True	True
PC FE	True	True	True	True

1995-05				
	(1)	(2)	(3)	(4)
$I(MSA\,Match)$	0.0119***	-0.0045***	0.0069***	0.0053***
	(0.0007)	(0.0007)	(0.0008)	(0.0009)
$sim_{LDA}(pc_i, pc_j)$	0.2459***		0.2423***	
	(0.0011)		(0.0012)	
$sim_{LDA}(pc_{i,MSA_i}, pc_{j,MSA_j})$		0.2651***		0.2743***
		(0.0012)		(0.0015)
$I_{MSA} * sim_{LDA}(pc_i, pc_j)$			0.0133***	
			(0.0019)	
$I_{MSA} * sim_{LDA}(pc_{i,MSA_i}, pc_{j,MSA_j})$				-0.0262***
				(0.0022)
N	438117	383569	438117	383569
Adjusted R^2	0.24	0.24	0.24	0.25
Year FE	True	True	True	True
PC FE	True	True	True	True

2005-15				
	(1)	(2)	(3)	(4)
$I(MSA\,Match)$	0.0079***	-0.0062***	0.0057***	0.0019**
	(0.0006)	(0.0007)	(0.0007)	(0.0008)
$sim_{LDA}(pc_i, pc_j)$	0.2430***		0.2416***	
	(0.0009)		(0.0010)	
$sim_{LDA}(pc_{i,MSA_i}, pc_{j,MSA_j})$		0.2573***		0.2637***
		(0.0010)		(0.0012)
$I_{MSA} * sim_{LDA}(pc_i, pc_j)$			0.0051***	
			(0.0016)	
$I_{MSA} * sim_{LDA}(pc_{i,MSA_i}, pc_{j,MSA_j})$				-0.0191***
				(0.0017)
N	569947	517333	569947	517333
Adjusted R^2	0.26	0.27	0.26	0.27
Year FE	True	True	True	True
PC FE	True	True	True	True

Table E.5: Regression results for strict local spillovers with raw data for LDAVec similarity.

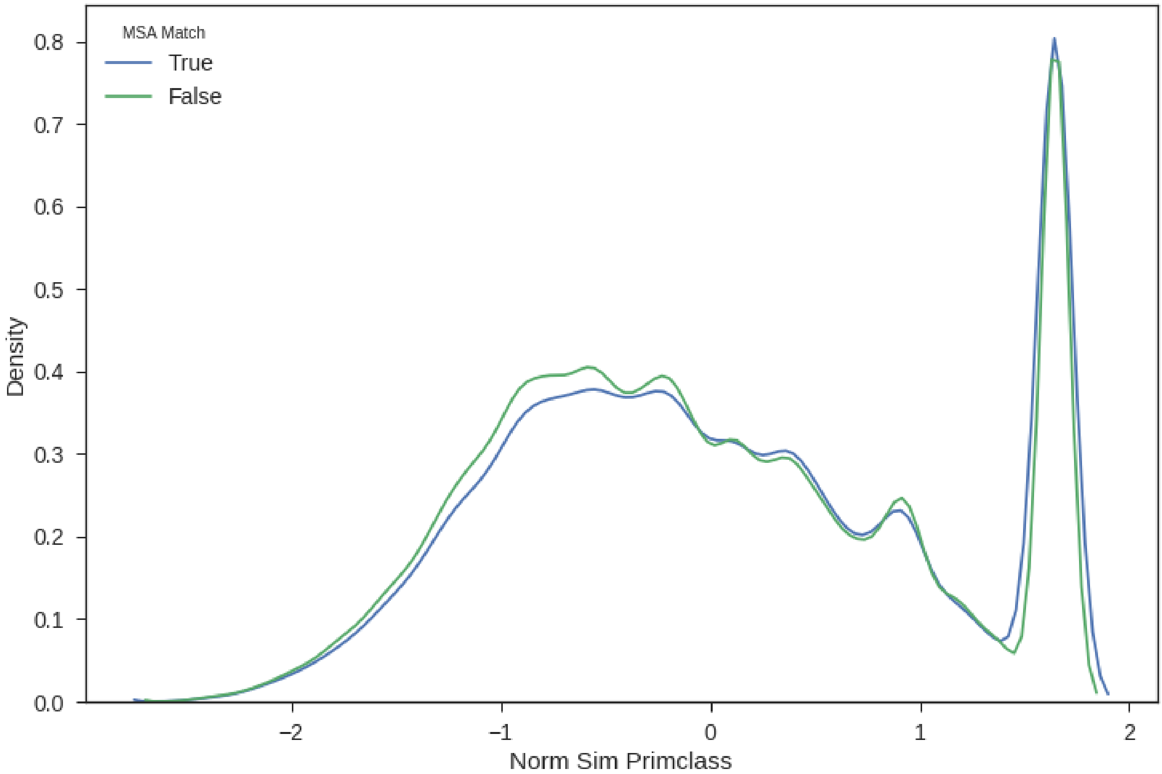


Figure E.1: Distribution of $sim(pc_i, pc_j)$ for local and non-local patent pairs.

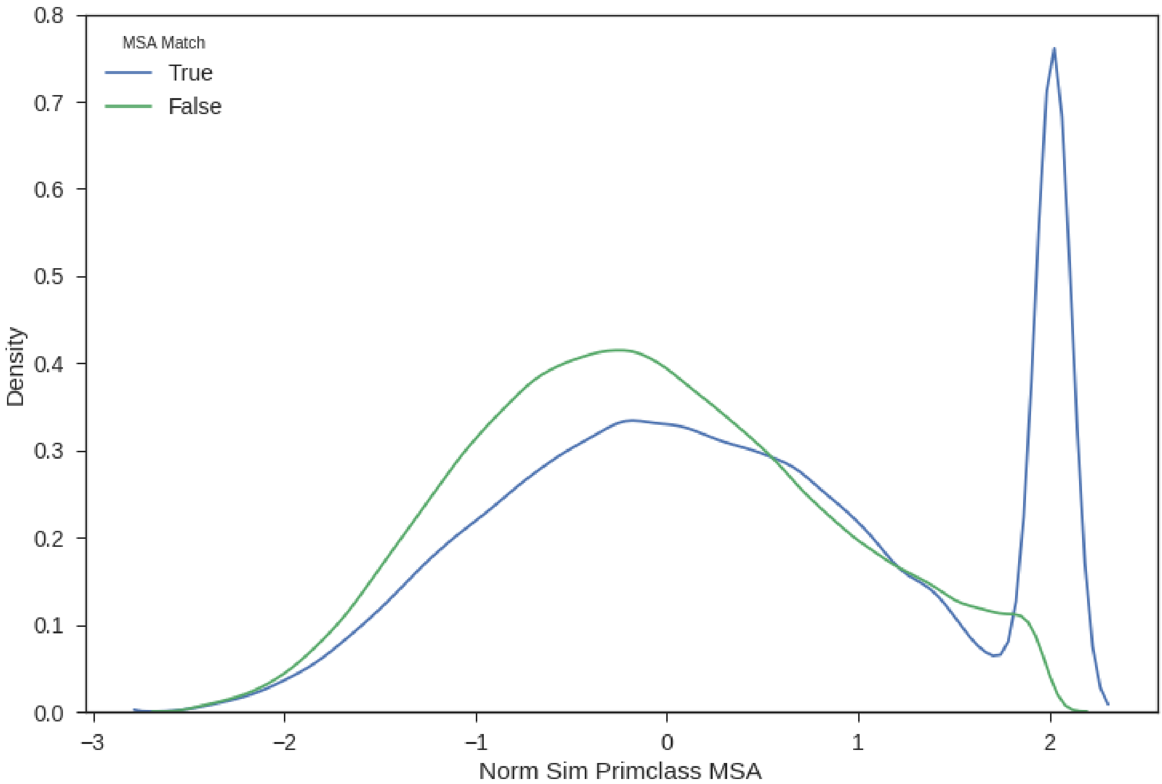


Figure E.2: Distribution of $sim(pc_i, pc_j)$ for local and non-local patent pairs.

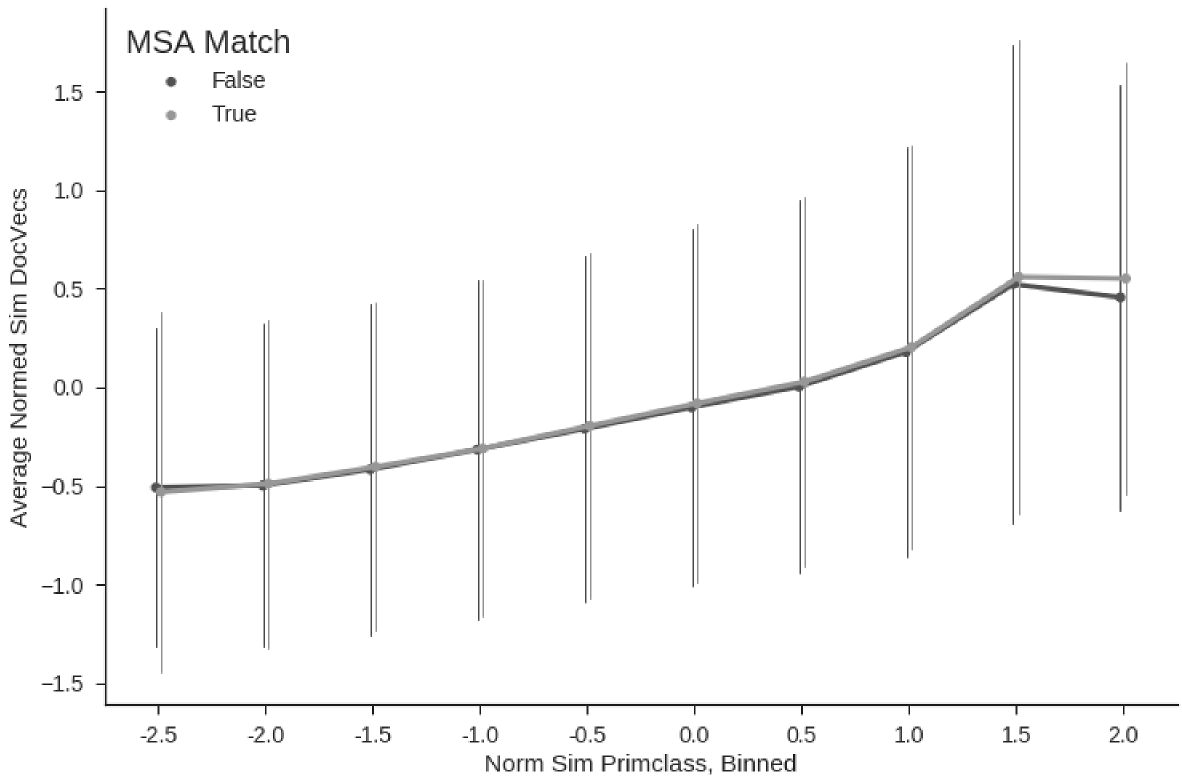


Figure E.3: Conditional means of DocVec similarity by primary class similarity.

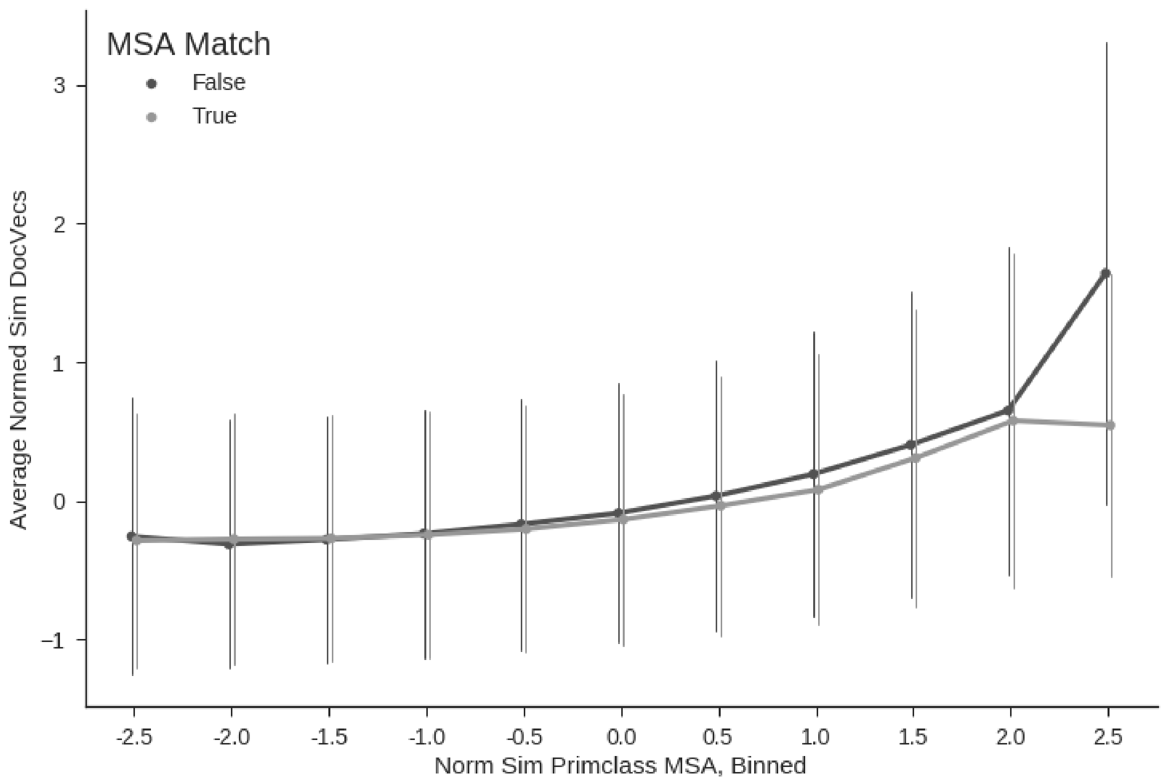


Figure E.4: Conditional means of DocVec similarity by MSA-primary class similarity.

F. Exogenous new knowledge

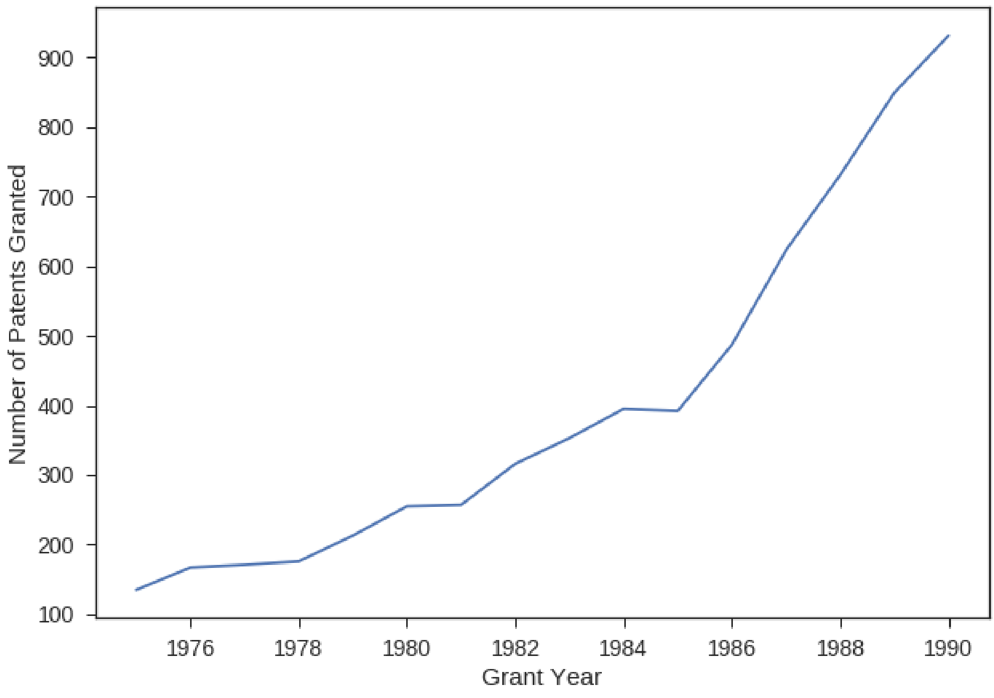


Figure F.1: Number of university patents by application year

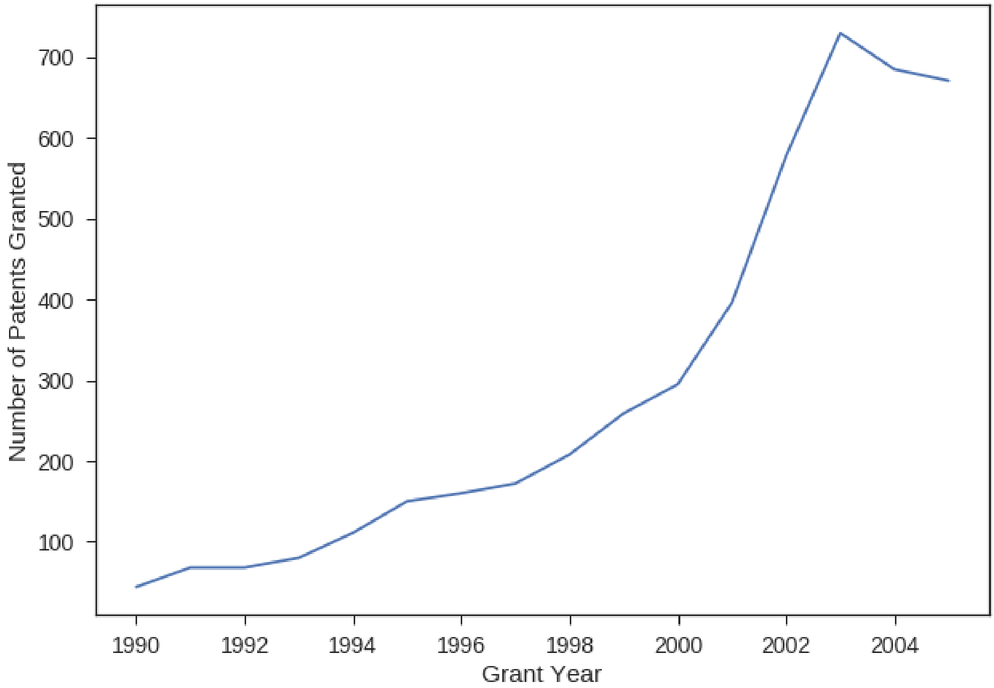


Figure F.2: Number of nanotechnology patents by application year

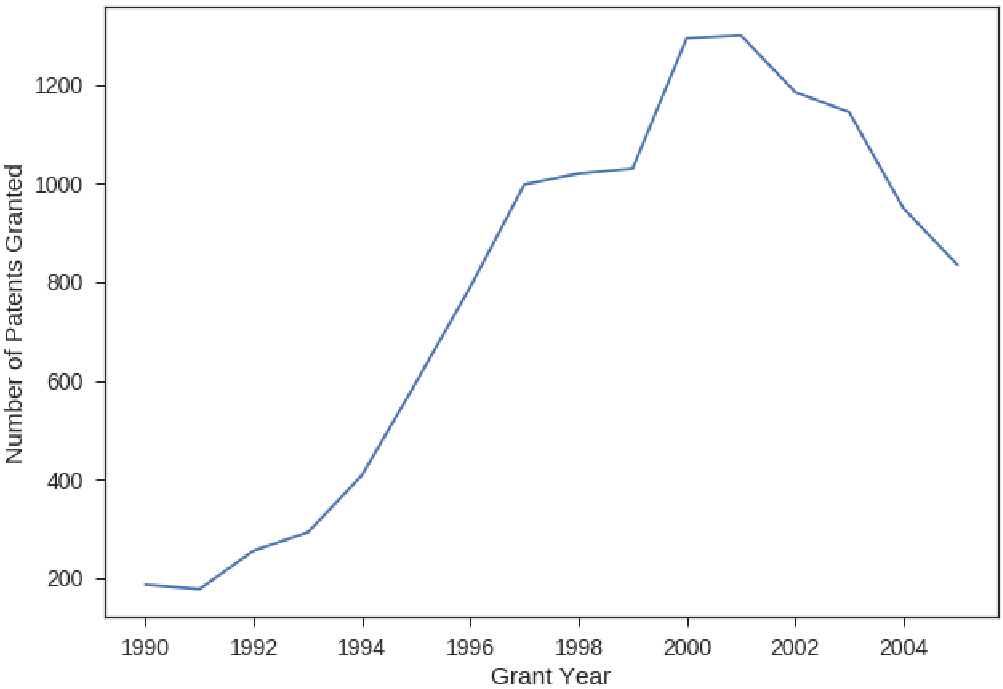


Figure F.3: Number of software patents by application year

G. New Terms

Some words at the end

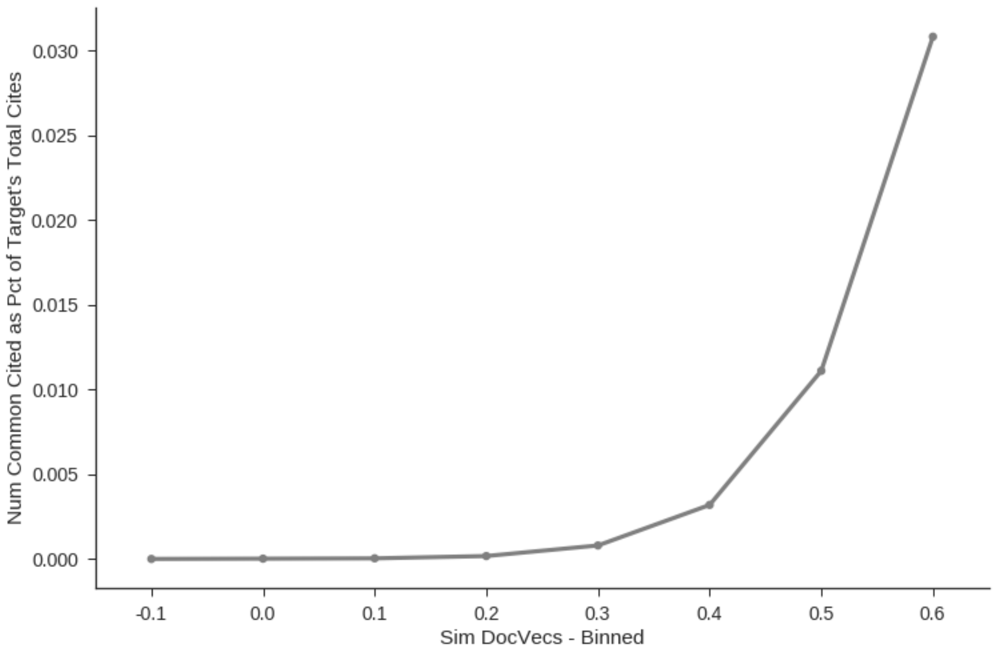


Figure G.1: Number of common cited patents as a percentage of target's total citations, by DocVecs similarity

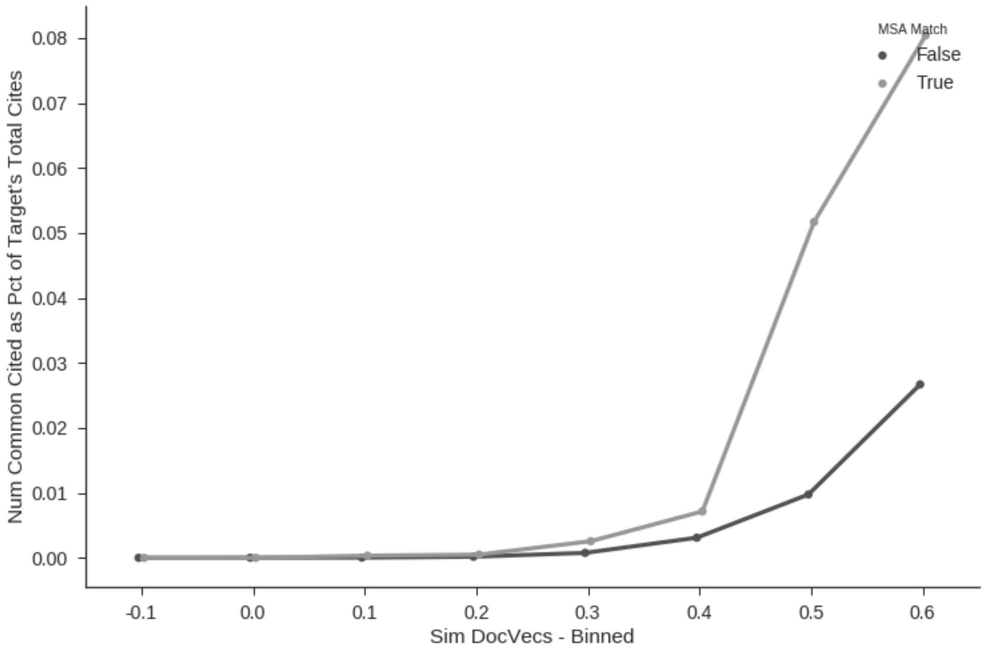
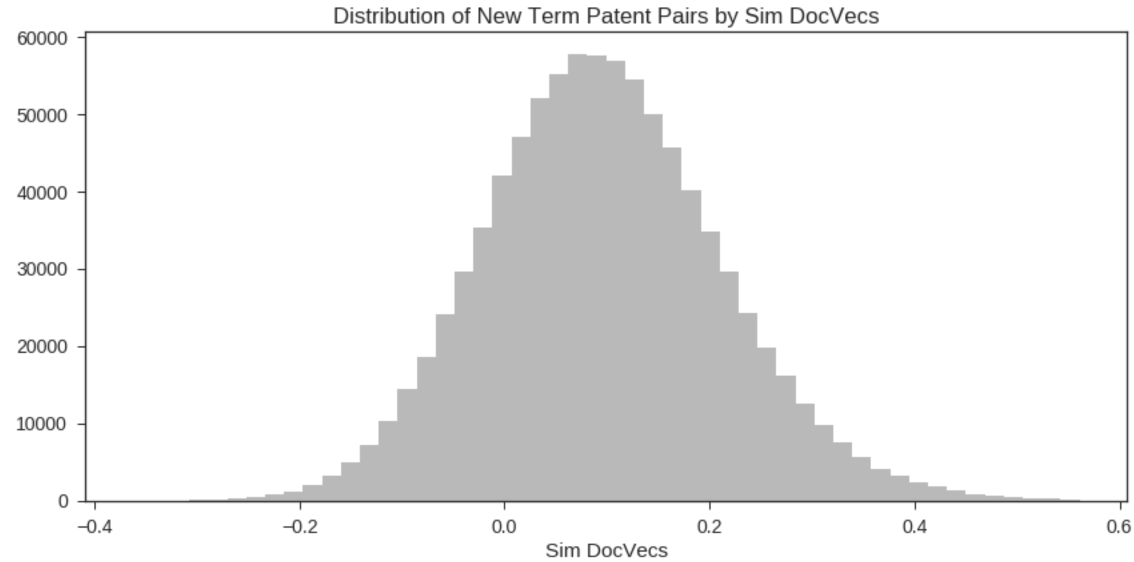
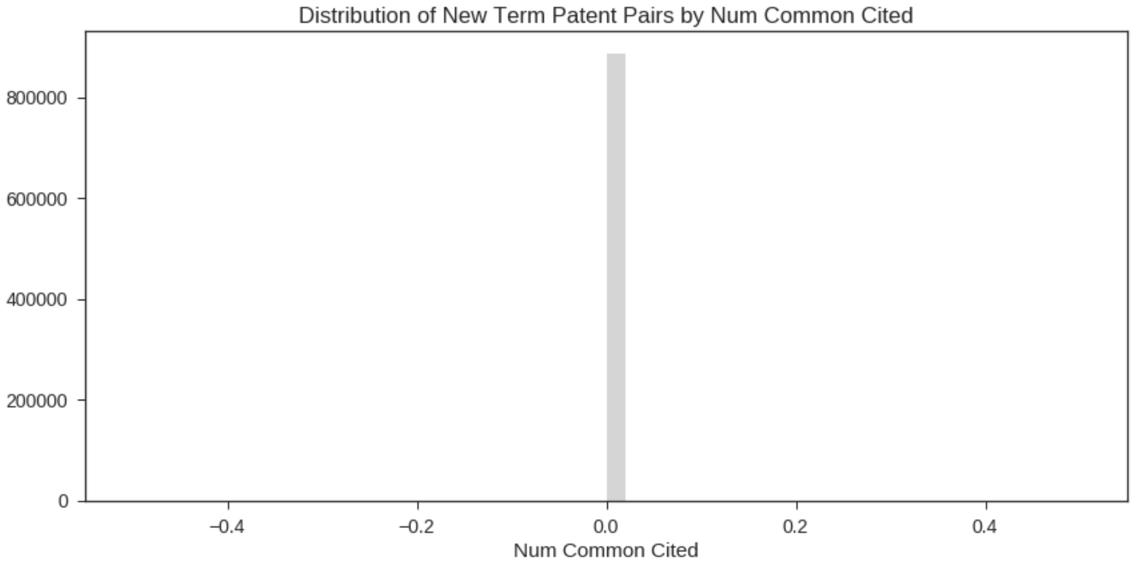


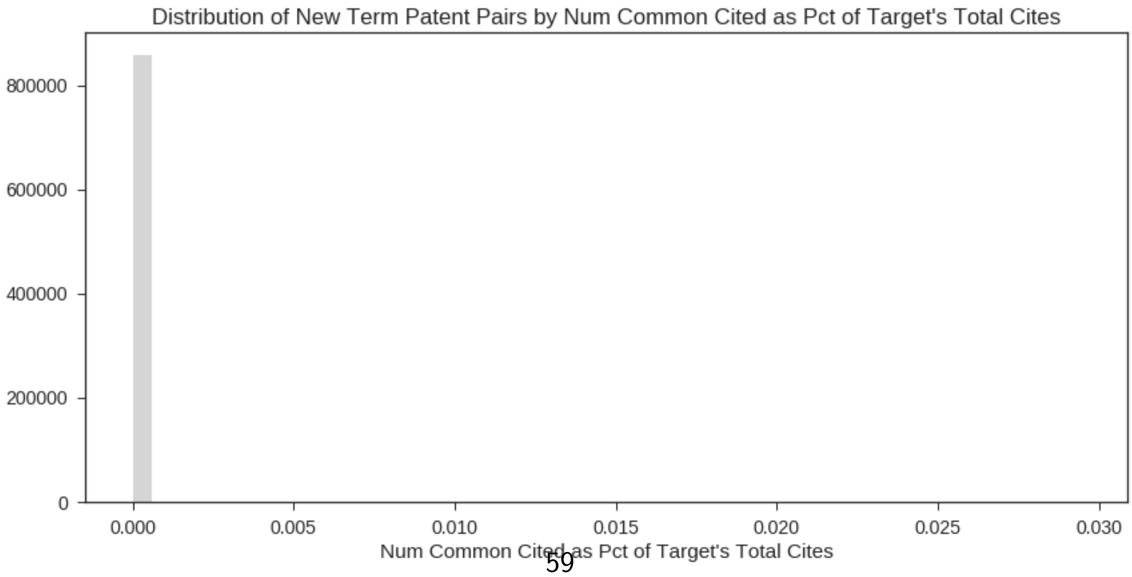
Figure G.2: Number of common cited patents as a percentage of target's total citations, by DocVecs similarity

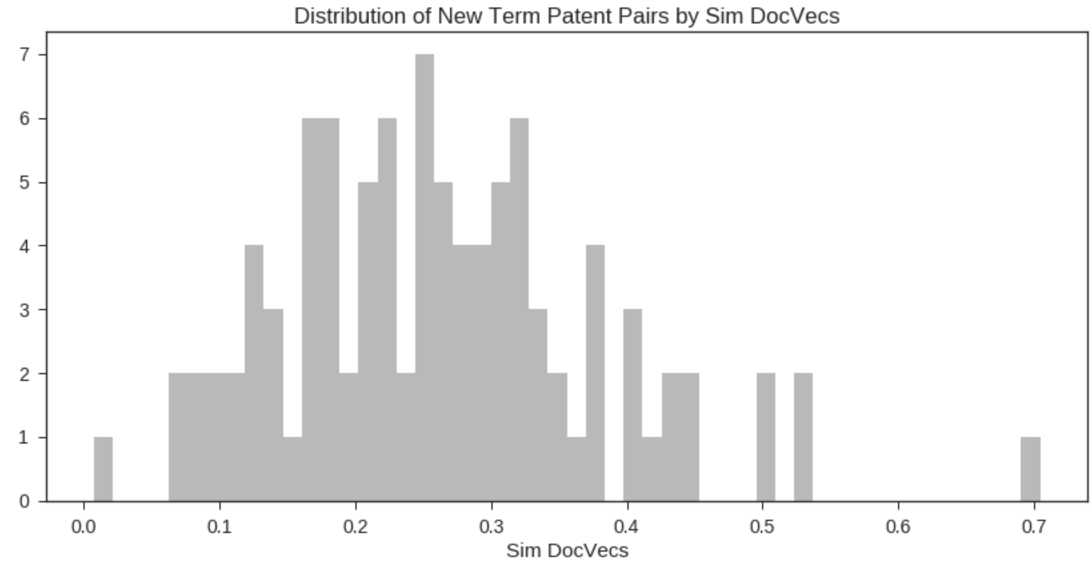
	Sim DocVecs	Num Common Cited	Pct Common Cited, Target's Citations	First Year	First Year, Num Pats	First Year, Num Pairs
gui	0.09	0.00	0.00	1992	1348	796727
lun	0.12	0.00	0.00	1995	415	63044
asic	0.11	0.00	0.00	1987	198	16716
url	0.10	0.01	0.00	1995	111	4847
serd	0.12	0.02	0.00	1998	75	1929
chat	0.10	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

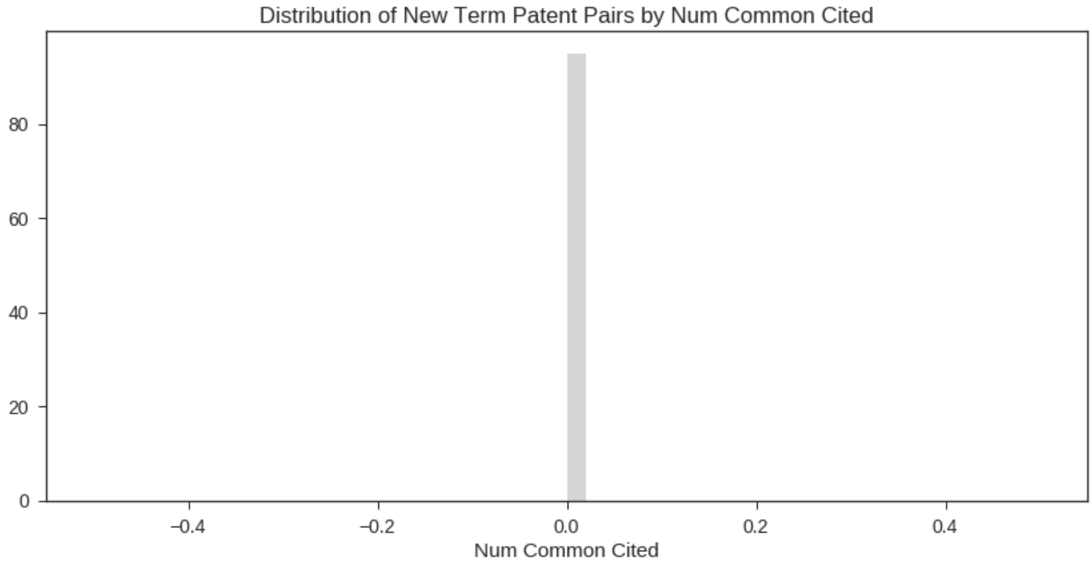
Table G.1: Comparison of similarity and backward citation overlap for new patents using new terms.

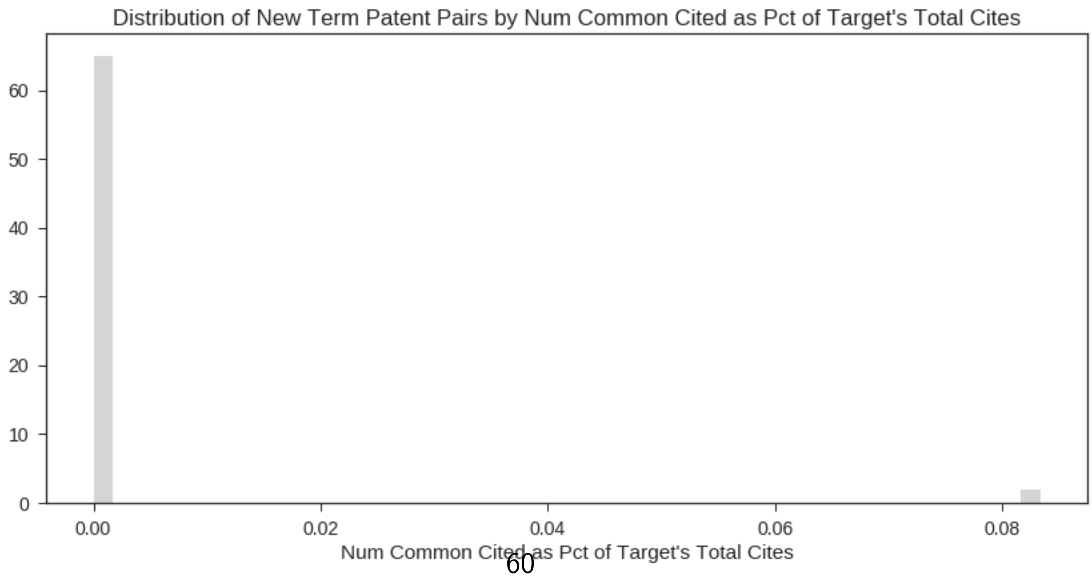

Distribution of New Term Patent Pairs by Sim DocVecs
60000
50000
40000
30000
20000
10000
0
−0.4
−0.2
0.0
0.2
0.4
0.6
Sim DocVecs


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
Sim DocVecs


Distribution of New Term Patent Pairs by Num Common Cited
Num Common Cited


Distribution of New Term Patent Pairs by Num Common Cited as Pct of Target's Total Cites
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 Wilbur Chung. Location strategies and knowledge spillovers. *Management science*, 53(5):760–776, 2007.

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.

Ashish Arora. Patents, licensing, and market structure in the chemical industry. *Research policy*, 26(4-5):391–403, 1997.

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.

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.

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.

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

Maryann P Feldman. Knowledge complementarity and innovation. *Small business economics*, 6(5):363–372, 1994.

Edward L Glaeser, Hedi D Kallal, Jose A Scheinkman, and Andrei Shleifer. Growth in cities. *Journal of political economy*, 100(6):1126–1152, 1992.

Michael Greenstone, Richard Hornbeck, and Enrico Moretti. Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings. *Journal of Political Economy*, 118(3):536–598, 2010.

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.

Jane Jacobs. *The economy of cities*. Vintage Books, 1969.

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.

Paul R Krugman. *Geography and trade*. MIT press, 1991.

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.

Sergey Lychagin, Joris Pinkse, Margaret E Slade, and John Van Reenen. Spillovers in space: does geography matter? *The Journal of Industrial Economics*, 64(2):295–335, 2016.

Charles F Manski. Economic analysis of social interactions. Technical report, National bureau of economic research, 2000.

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

Alfred Marshall and Mary Paley Marshall. *The economics of industry*. Macmillan and Company, 1920.

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. Do patents weaken the localization of innovations? evidence from world's fairs. *The Journal of Economic History*, 71(2):363–382, 2011.

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.

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.

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