Patent Application Citations and the Examination Process: A Network-Based Date-Sensitive Method to Analyze Patent Applications

Ryan Whalen
Northwestern University
2240 Campus Drive
Evanston Il, 60208
+1 773.679.1344
r-whalen@northwestern.edu

Noshir Contractor
Northwestern University
2240 Campus Drive
Evanston Il, 60208
nosh@northwestern.edu

ABSTRACT
This paper enriches the patent application data available from the USPTO by extracting citations from the application full text. We subsequently use this data to perform a novel boundary spanning analysis, scoring applications based on the degree to which they span technologically-distant areas at the time the application is filed. Finally, we show that while including citations in patent applications is associated with a higher probability the application will be granted, too many citations can have the opposite effect. Similarly, citing to disparate technological areas increases the time required by the USPTO to assess an application, and is correlated with a lower probability the application will be granted.

Categories and Subject Descriptors
K.5.1 [**Legal Aspects of Computing**]: *patents, regulation*

J.4 [**Social and Behavioral Sciences**] *economics, sociology*

General Terms
Algorithms, Management, Measurement, Economics, Legal Aspects.

Keywords
Patent Citations, Patent Applications, Knowledge Flow, Innovation, Patent Examination, USPTO

1. INTRODUCTION
Increasing access to both legal data and computational power is transforming empirical legal research. As data access and analysis capabilities have improved, researchers have moved towards analyzing ever-newer-and-larger sets of legal data. In order to enable this work, researchers require new datasets and methods. This paper addresses both of these requirements by adding value to an existing legal dataset—U.S. patent application data—and by developing a new citation-network-based method of measuring technological boundary spanning.

Patenting generates an enormous amount of data. This data can provide valuable insight into the relationships between patent granting agencies, patent law and innovation. The data arising from granted patents has been extensively used by scholars of law, economics and management [9,19,31]. However, the data arising from patent *applications* has been much less frequently used, even though it provides insight into a particularly important part of the patenting process: the examination.

This paper begins to address this lack of focus on patent application data and the citations they make. We create a new dataset that tracks all citations from U.S. patent applications to granted U.S. patents, and develop and test a method to measure the degree to which any application spans disparate technological disciplines.

We show that the citations included in patent applications are important to the examination process. Including citations in an application is correlated with a higher probability that the application will be granted by the USPTO. However, if the application cites uncommonly combined technological precedent, the probability of granting decreases while the examination time increases.

The dataset this paper assembles, and the new method for measuring the degree to which an application spans technological boundaries, will both prove useful to scholars of patent law, and innovation policy.

2. Citations
Citations have long provided valuable information about the legal system's structure. Since the 19th century development of legal citators like *Shepard's Citations* [20], lawyers have long thought of the law as linked together into a large citation network [21,28]. Recent developments in computational power and network analysis techniques have enabled researchers to analyze these legal networks at large scale and in novel ways [10,11,33]. In recent decades, patent citations have joined precedent citations as a valuable source of legal data.

2.1 Patent Citations
Patent citations lay bare the knowledge network that underlies our innovation system. They can be viewed from two perspectives: By looking at the backward citations that patents and patent applications make, we can situate those inventions in the knowledge network; alternately by looking at forward citations we can generate impact scores and proxy measures of patent value. While there are few studies examining patent *application* citations, the same is not true of granted patent citations

2.1.1 Granted Patent Citations
The citations in granted patents have long provided a fruitful source of data for researchers interested in patent law, knowledge flows, and collaboration. Early proposals for a citation system to track the relationships between patented inventions hoped to help searchers more readily determine the state of the art [26], and

track relationships between technology classes [15]. Since the USPTO began including patent citation information in patent disclosure documents, researchers have used both the forward and backward citations to gain insight into the innovation system.

2.1.1.1 Forward Citations.

Researchers have long used forward patent citations as proxy measures for an invention's "technological significance" [23:167] or value [14,31]. Studies have shown that highly-cited patents are more likely to receive industry recognition [5], are more likely to be renewed by their owners [14,22], and are more likely to be identified as important by experts [1].

In addition to their use as proxy measures for an invention's significance or value, researchers have used aggregate patent citations to assess national innovation policy [16,18] and firm performance [13]. These aggregate citation studies have demonstrated how patent citations can be useful in assessing a country's dependence on foreign research [18], shifts in the geography of innovation [16], and how effective a firm's R&D investments are [13,31].

Along with use as a proxy measure for technological value or significance, patent citations can also help trace the flow of knowledge. This work has shown that patents are much more likely to cite to other patents from the same geographic area [17,29], and that they are also more likely to rely on local science when citing to non-patent prior art [30].

2.1.1.2 Backward Citations.

In addition to in-coming or forward citations—that is the citations that a patent receives from other later inventions—researchers have also used out-going or backward citations in order to learn more about the citing patent. For instance, research shows that patents citing across disciplines tend to have greater impact [27]. One of the theoretical explanations for this phenomenon is that boundary-spanning patents may draw together knowledge in new ways, generating more novel inventions.

Backward citations can also provide a useful measure of knowledge flow. In doing so, citations are taken as evidence of the flow of knowledge between individuals and firms [See e.g., 3,12,24]. This body of work assumes that citations from one patent to another represent knowledge that has transferred from the inventors or firms listed on the cited patent to those on the citing patent—whether that transfer occur by teaching, otherwise sharing knowledge, or via the disclosure function of the patent system. However, measuring these knowledge flows is complicated by the fact that many of the citations in granted patents are placed there by examiners rather than by the applicants [2].

In addition to mapping past flows of knowledge, researchers have used patent citations to predict future evolution of the invention–information network [8]. By clustering patents based on class-to-class citation patterns, Érdi et. al demonstrate the potential to forecast the emergence of new technological fields as inventions of one class begin to draw on inventions of another creating a novel recombination.

2.1.2 Patent Application Citations.

While granted patent citations have provided a rich source of data for scholars of innovation, there are few studies analyzing the citation data in patent *applications*.[1] Since 2001 the USPTO has published the full text of the vast majority of patent applications. These applications provide a valuable source of data showing the patterns of knowledge flow and information relations amongst inventions that both eventually succeed in passing muster at the USPTO and those that fail. The ability to observe both successful and unsuccessful applications enables us to model predictors of patenting success.

Those few studies that have examined citations appearing in patent applications have limited their scope to those applications that go on to be granted. Sampat [25] shows that the tendency to cite prior art at the application stage varies by industry and reflects the strategic motivations for seeking patent protection. Industries that are more likely to litigate their patents are also more likely to include prior art citations in their applications because doing so can help insulate them against future legal attacks on their patent rights. In addition, applicants are more likely to include citations in applications for valuable inventions—also presumably in order to help insulate them against future legal attacks.

While patent applicant citations may appear to be strategically motivated, there is evidence suggesting that USPTO examiners do not always take applicant-provided citations into account when rejecting applications—preferring instead references they find during their own prior art searches [7]. This somewhat complicates the interpretation of applicant-provided citations. If examiners truly are "myopic" in the execution of their duties, preferring prior art they discover themselves over that provided by outside sources, then it may be that applicant-provided citations have little relationship to whether or not an application is eventually granted.

So, despite the fact that granted patent citations have provided a rich source of research data, patent application citations have gone relatively underutilized. Those few studies that have examined them have focused either on the applicant-included citations as present in final patent grants, or on applications that go on to be narrowed by the USPTO. To date, no study has systematically examined the relationship between citations included in patent applications and how the applications fare in the patent examination process. This article is the first to assemble and analyze the citation data in both granted and ungranted patent applications. This rich source of data raises a number of research questions and hypotheses.

3. Questions & Hypotheses

3.1 Grant Likelihood and Motivations for Citing Prior Art

There are a variety of plausible reasons to include prior art citations in a patent application. Most of these reasons lead us to believe that citations will positively correlate with grant probability.

The most obvious reason for including patent citations in an application is to comply with patent law. Patent law imposes a duty upon applicants to disclose "material" prior art. While there is no duty to search for relevant prior art to disclose, if applicants are aware of any prior art that would establish—either on its own or in combination with other references—a prima facie case of

[1] This is likely due to the fact that the USPTO's publicly available bulk data provides machine readable citation fields for granted patents, but not for patent applications.

unpatentability, they are obliged to refer examiners to that prior art. Given this duty to disclose relevant prior art, we would expect that better-informed applicants would be more likely to include citations. These better-informed inventors will be aware of more of the relevant prior art, leaving them more likely to be obliged to disclose references.

Citing prior art is also advantageous because it can strengthen a patent against later legal attacks. If, after the patent is granted, opponents point to prior art in an attempt to invalidate the patent, the presumption of validity is stronger if the patent examiner considered that prior art during the examination process.

In addition to the legal duty to cite prior art, applicants may make citations for strategic reasons. They may wish to assist the patent examiner in performing her prior art search, and thereby increase the likelihood that the applicants will be granted a patent. Alternately, applicants may be attempting to guide the prior art search in a particular direction, again hoping to influence the ultimate granting decision.

These motivations for including prior art in a patent application—demonstrating the applicant's knowledge, strengthening the patent against later attack, and guiding the prior art search—all suggest that including citations will be positively correlated with an application's probability of being granted. This leads to our first hypothesis:

H1: Patent applications that include citations to previously patented inventions will be more likely to be granted patents by the USPTO.

3.2 Boundary spanning patent applications

Along with demonstrating applicant knowledge, defending against future challenges, and facilitating examiner prior art searches, citations in patent applications also demonstrate the sources of knowledge underpinning the inventions that applicants seek to patent. This information can provide insight into how any given application fits into the patent citation network; whether they are embedded in local technological areas or whether they span technological boundaries.

Previous research has shown that granted patents featuring citations across technological domains are more successful [27]. Similarly, successful scientific publications also tend to feature an atypical combination of cited sources [32]. This could suggest that inventions crossing disciplinary boundaries may be more likely to be successful.

On the other hand, citations that span technological boundaries may suggest that an application is attempting to claim a more complex and unlikely invention. In addition, by bridging multiple areas of prior art, citations from one technology class to another may complicate an examiner's job by expanding the universe of prior art she needs to search.

These divergent possibilities—that boundary crossing citations may alternately increase or decrease the probability of success—suggest the following question:

RQ1: Are patent applications that cite across disciplinary boundaries more or less likely to be granted by the USPTO?

3.3 Boundary spanning applications and examination complexity.

While citing across disciplinary boundaries may alter the probability of successfully attaining a patent, it also increases the technological space that one needs to be familiar with in order to determine whether an invention is useful, novel and nonobvious as required by the USPTO. Patent examiners are required to perform a thorough prior art search to determine whether a patent application contains claims that fulfill the patentability requirements. The more technological areas implicated by an invention, and the more unlikely their combination, the more involved their search must be, thus:

H2: Patents citing across disciplinary boundaries will have longer pendency periods

3.4 Team size and boundary spanning applications.

In addition to taking more time to assess, inventions that draw upon diverse sources of knowledge require more diverse expertise to invent in the first place. As science and technology have increased in complexity, the popularity of team science and the size of teams have also increased [34]. These teams are more likely to draw on atypical combinations of sources [32]. If this holds true for inventors at the patent application stage we would expect to see that:

H3: Patent applications with more than one listed inventor are more likely to cite across disciplinary boundaries.

In order to test these hypotheses, we assemble a unique dataset containing the citations to United States patents in patent applications filed between 2001 and 2006, and subsequently use this data to examine their relationship with the USPTO examination process.

4. Methods

4.1 The Data

Unlike the granted patent data, patent application data does not include a machine-readable field listing the application's citations. Any citations made by the applicant are included in the full text of the application, generally in the description field. In order to extract this data we performed pattern matching via a regular expression coded to match citations to United States patents.

Because a number of the above hypotheses rely on testing whether or not a given application was ultimately granted, only patent applications published between the beginning of 2001 and the end of 2006 are included in the analysis.[2] Limiting the set of applications under analysis to those filed by the beginning of 2006 allows time to account for the pendency period as patents are under examination. As of 2006, the average patent application spent 22.6 months in pendency [6], so focusing on pre-2006 applications provides sufficient time for the majority of the analyzed applications to work their way through the examination process.

Along with citation information, we extracted a number of additional variables from the XML files published by the USPTO.

4.2 Boundary Spanning.

In order to measure the degree to which an application spans technological boundaries we rely upon the technology classes assigned to the patent applications and the patents they cite. We know that patents are likely to cite other patents within the same

[2] The USPTO did not begin regularly publishing application data until 2001. While we limit the data analyzed here to allow time for application pendency, we have extracted citation data for patent applications through 2010, and the publicly available dataset includes all of these applications.

class [27]. This suggests that when they cite to another technology class they are, in a sense, spanning technological boundaries.

The challenge in measuring boundary spanning arises when trying to assess the degree of boundary spanning any given citation might represent. Technology classes have varied intra and inter-class citation rates, so simply treating citations as binary same-class or other-class variables would not accurately reflect the novelty of any given citation. After all, if class A almost always only cites other class A inventions, then a cite to class B is truly notable as a boundary spanner. However, if class A has often cited class B in the past, then another A-to-B citation would be less noteworthy.

In order to address these inter-class differences in insularity we need to account for varied rates of citation between classes. In addition, because there is likely to be changes in citation practices as technologies evolve [4], we also need to account for changes over time. Accordingly, we propose a network-based date-sensitive method to assess patent citation boundary spanning. This involves a multistep process to calculate a boundary spanning score for each observed citation from a patent application to a U.S. patent. This score is defined as:

$$1 - \left(\frac{\sum n_j}{\sum n} \times \frac{m_{ij}}{\left(\frac{\sum n_j}{\sum n} \times m_{i\_} \right)} \right)$$

Where m_{ij} is the number of citations between the source class i and the target class j, $m_{i\_}$ is the total number of citations from class i, n is all patents, and n_j is patents of the cited class j.

This calculation takes into account the citations we would expect to see between two classes given the distribution of patent classes, and adjusts those expectations given previous citation patterns from one class to another. Because these adjustments depend on the historical citations made before the patent application is filed, they vary depending on the filing date.

In order to address this variation we process the patent applications in order on a day-by-day basis. We begin with the full utility patent citation network for all patents issued from the beginning of 1975 through until the day preceding the application filing date. We then examine each application filed on that date, and each citation they make. Noting the class of the application and the class of the patent it is citing we then calculate how atypical that citation is. For clarity we can break calculation of the above formula down into a step-by-step process:

Step 1: Calculate the naive citation probability, representing the likelihood we would expect to see a citation from class i to class j if citations were distributed at random: $a = \frac{\sum n_j}{\sum n}$.

Step 2: Calculate the total expected citations we would expect to see from class i to class j given how many citations there have been from class i from 1975 up until the day of filing: $\omega = a \times m_{i\_}$.

Step 3: Taking into account the observed number of citations and the expected number of citations, calculate the over/under performance of citations from class i to class j: $\theta = \frac{m_{ij}}{\omega}$.

Step 4: Calculate a boundary spanning score weighting the naive probability by the over/underperformance. Subtract it from 1 so that uncommon citations score highly and more predictable citations score lowly: $1 - a \times \theta$.

Step 5: After scoring citations for all applications made on day t, update the citation network to include patents granted and citations made up until day t+1. For each citation filed on day t+1, repeat steps 1–4.

4.3 Control variables.

Team size is a significant predictor of scientific success [34] and teams are more likely to rely on atypical combinations of sources [32]. Given the importance of team size on both citation practices and outcome, when assessing these factors it is necessary to control for the number of inventors. This variable was computed by simply counting the number of inventors listed on each patent application.

In addition to team size, we also calculated the number of figures or drawings included with the patent application, the number of independent claims made by the application, the USPTO main technology class assigned to the application, the filing date, and whether the application claimed foreign priority.

5. Results

The dataset includes all utility patent applications submitted between January 1, 2001 and December 31, 2005 and subsequently published by the USPTO. This end date was chosen to provide sufficient time for applications to work their way through the examination process, so that we may determine whether they were ultimately granted. For each application, the *granted* variable was coded as 1 if any utility patent issued before 2012 listed that application's number in the application number field.

5.1 Descriptives

The final dataset includes data on 1,385,619 utility applications filed during this time period. Of these, 461,412 had citations to patents filed after 1975, with a total of 4,256,535 citations.[3]

Descriptives	
Applications	1,385,619
Mean citations / application	3.09 (s.d. 25.54)
Applications with Citations	461,412
Mean citations / application with > 0 citations	9.22 (s.d. 43.72)

Because the application citation data is computationally extracted, we validated our method by manually comparing the results of our citation extraction program against the patent application text as published in the USPTO PAIR database. We manually checked hundreds of randomly selected citations and dozens of randomly selected applications and in each instance our citation extraction method captured all citations to United States patents and in no instance were citations added where they did not exist. To our knowledge this is the first dataset of its kind that includes patent

[3] Because we require the utility patent citation network details to calculate boundary spanning, only citations to post-1975 patents are used in the following analyses. This excludes a small portion of citations to older patents (approximately 5.8% of the observed application citations were to pre-1975 patents).

application citation data for both granted and ungranted patent applications.[4]

5.2 Citations and Grant Rates

There are a number of reasons to believe that including citations in an application will improve the odds that a patent is ultimately granted. Citations to relevant prior art may make it easier for overworked examiners to assess an application. The citations give them somewhere to begin their prior art search and perhaps some assurance that the applicant is familiar with the field. Applicants also have a duty to disclose relevant prior art that they are aware of, and because they enjoy a stronger presumption of validity if the prior art is considered by the examiner, they have an interest in presenting it for patents that they anticipate will be valuable and potentially litigated. Table 1 shows the results of two logistic regression models with each application's ultimate success or failure at the USPTO as the dependent variable (granted = 1, not granted = 0). The first column shows that including citations does indeed significantly increase the probability that an application will be granted. The baseline citation value in this first model is 0, and we can see that any number of citations significantly increases the likelihood that an application will be granted. The model controls for the main technology class that the USPTO assigns to each application.

The second model excludes applications that did not include any citations, so the baseline number of citations here is 1. Here we see that, for the most part, including more citations does not increase an application's odds of success. In fact, there is a small but statistically significant negative effect for applications that include more than 8 citations. This finding that including citations improves the odds that an application will be granted a patent supports H1.

Table 1[5]

D.V.: *Granted*	All Applications	Applications With Citations
1 Citation	0.092 ***	
	(0.006)	
2–3 Citations	0.095 ***	0.006
	(0.006)	(0.009)
4–8 Citations	0.101 ***	0.012
	(0.007)	(0.009)
> 8 Citations	0.032 ***	-0.024 *
	(0.007)	(0.01)
Inventors	0.05 ***	0.037 ***
	(0.0004)	(0.002)
Independent Claims	0.005 ***	0.003 ***
	(0.0004)	(0.0004)
Figures	0.00009	-0.01 ***
	(0.00008)	(0.0001)
Fixed Class	X	X
	N = 1,382,512	N = 457,675
Signif. codes:	0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1	

5.3 Boundary Spanning and Grant Rates

Atypical combinations of citations may suggest that an application is attempting to patent an invention that is more likely to be novel and nonobvious—the two most important patentability considerations. This could lead applications featuring atypical citations to be more likely to be granted. Alternately, atypical citations could correlate with inventions that are more complex and difficult to perfect. They also expand the universe of prior art, making examiners' jobs that much more difficult. Both of these factors could decrease grant probability.

Table 2 demonstrates that the latter is the case. Applications that feature citations that are unlikely given the distribution of citable patents, and citation history are less likely to be granted.

[4] Researchers interested in replicating our results or using the application citation dataset for their own purposes can contact the author for access.

[5] The citation intervals (e.g. 1, 2–3, 4–8, > 8) were determined by splitting at the citation quartiles for all applications that had at least one citation. So there is approximately the same number of applications in each citation group. The values in parentheses are standard errors.

Table 2

D.V.: *Granted*	Applications With Citations
Atypical Citations	-0.456 ***
	(0.016)
2–3 Citations	0.012
	(0.009)
4–8 Citations	0.025 **
	(0.009)
> 8 Citations	-0.0003
	(0.01)
Inventors	0.038 ***
	(0.0016)
Independent Claims	0.003 ***
	(0.0004)
Figures	-0.0009 ***
	(0.0001)
Within Class	X
	N = 457,675

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

5.4 Boundary Spanning Citations and Patent Pendency

H2 suggests that, because they will be present in more complex inventions, and because they expand the universe of prior art that an examiner must consider before granting a patent, applications featuring atypical citations will have longer pendency periods. Table 3 reports the results of OLS regression models for both all granted applications and those with citations. The second model supports H2 demonstrating that boundary spanning citations are significantly related to increased examination times. One standard deviation reduction in the boundary spanning measure results in 40 fewer days in pendency. It is also interesting to note that including citations in applications reduces the pendency period. This could be because they assist the examiner in her prior art search.

Table 3

D.V.: *Pendency* (days)	Granted Applications	Granted Applications With Citations
Atypical Citations		164.10 ***
		(4.53)
Inventors	9.05 ***	10.12 ***
	(0.29)	(0.48)
Figures	0.76 ***	0.83 ***
	(0.03)	(0.07)
Independent Claims	2.45 ***	1.12 ***
	(0.07)	(0.08)
2–3 Citations	-18.66 ***	-23.07 ***
	(1.80)	(2.49)
4–8 Citations	-38.84 ***	-45.21 ***
	(1.86)	(2.63)
> 8 Citations	-57.82 ***	-45.41 ***
	(2.03)	(3.00)
Within Class	X	X
	N = 846,604	N = 272,572
	Adj R-squared 0.25	Adj. R-squared 0.18

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

5.5 Collaboration and Atypical Citations

H3 suggests that as the number of inventors increases, the likelihood that an application will feature a boundary spanning citation will also increase. The more inventors involved in the research and development process, the more likely they are to make citations across disciplinary boundaries. In addition, larger teams are likely to be familiar with more prior art than smaller teams, and thus, in fulfilling their duty to disclose the relevant prior art that they know about, they are likely to cite more sources. Table 4 supports H3, showing that an increased number of inventors is significantly correlated with high boundary spanning citation scores.

Table 4

D.V.: *Atypical Citations*	Applications With Citations
Inventors	0.53 ***
	(0.0057)
2–3 Citations	0.002 ***
	(0.0001)
4–8 Citations	0.032 ***
	(0.0009)
> 8 Citations	0.06 ***
	(0.0009)
Independent Claims	0.0002 ***
	(0.00003)
Figures	0.0001 ***
	(0.000008)
Within Class	X
	N = 457,675
	Adj. R-squared 0.26
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1	

6. Discussion

It is unsurprising that including citations improves one's chances of filing a successful patent application. This makes sense when considered both from the applicant and the examiner perspectives. Applicants with a greater knowledge of the prior art are more likely to be legally obliged to include citations. These are also the applicants who are most likely to know that their own application is indeed novel and nonobvious. Those applicants who do not include any citations are perhaps ignorant of the prior art and potentially unaware of references that would invalidate their claims.

Similarly, when considered from the examiner's perspective, it makes sense that applications with citations would have better outcomes. The most complex aspect of the examiner's job is the prior art search. Including citations helps the examiner discover the relevant prior art, making her job easier and perhaps assuring her that the applicant is knowledgeable.

The negative effect that citing across disciplinary boundaries has on the probability that an application will be granted can also be understood from both the applicant and examiner perspectives. From the applicant perspective, the inclusion of boundary spanning citations suggests an invention that is more complex difficult to perfect. Meanwhile, when considered from the examiner perspective, boundary spanning citations are likely to lead to a more time consuming and difficult prior art search. This may lead the famously overworked USPTO examiners to be more likely to reject the application.

This project is an early step in a larger project aimed at developing new methods to analyze and understand the vast amounts of data that the patent system produces. The application citation dataset described here will hopefully prove a valuable source of data to economists and legal scholarly. Similarly, we hope that our network-based method and date-sensitive boundary spanning measure provides a more nuanced and realistic view of the degree to which applications incorporate diverse technological references.

Going forward, more work is required to understand the causal mechanisms that lead boundary spanning applications to both take longer in assessment and ultimately be more likely to be rejected by the USPTO. It could be that they correlate with lower-quality inventions. Alternately, it could be that the USPTO is overworked and unable to adequately examine these unique sorts of interdisciplinary inventions. Without further research it is impossible to distinguish between these competing—or perhaps complimentary—explanations.

7. Conclusion

We contribute to the development of big data analysis in the law by developing a novel dataset and a new network-based and date-sensitive method for measuring the degree to which patent citations span technological boundaries.

Our analysis demonstrates that including citations increases the probability that a patent application will be granted, while citing across uncommonly spanned disciplinary boundaries decreases that probability. It also shows that teams are more likely to span boundaries, and when they do their applications are likely to take longer for the USPTO to process.

Future research will be able to use our data and methods to further explore these findings and further questions of interest to scholars of law and innovation.

8. REFERENCES

1. Albert, M.B., Avery, D., Narin, F., and McAllister, P. Direct validation of citation counts as indicators of industrially important patents. *Research policy 20*, 3 (1991), 251–259.

2. Alcácer, J. and Gittelman, M. Patent Citations as a Measure of Knowledge Flows: The Influence of Examiner Citations. *Review of Economics and Statistics 88*, 4 (2006), 774–779.

3. Almeida, P. and Kogut, B. Localization of Knowledge and the Mobility of Engineers in Regional Networks. *Management Science 45*, 7 (1999), 905–917.

4. Boyack, K.W., Börner, K., and Klavans, R. Mapping the structure and evolution of chemistry research. *Scientometrics 79*, 1 (2009), 45–60.

5. Carpenter, M.P., Narin, F., and Woolf, P. Citation rates to technologically important patents. *World Patent Information 3*, 4 (1981), 160–163.

6. Chung, J.J. Patent Pendency Problems and Possible Solutions to Reducing Patent Pendency at the United States Patent and Trademark Office. *J. Pat. & Trademark Off. Soc'y 90*, (2008), 58.

7. Cotropia, C.A., Lemley, M.A., and Sampat, B.N. Do Applicant Patent Citations Matter? *Research Policy 44*, (2013), 844–854.

8. Érdi, P., Makovi, K., Somogyvári, Z., et al. Prediction of emerging technologies based on analysis of the US patent citation network. *Scientometrics 95*, 1 (2013), 225–242.

9. Fleming, L. and Sorenson, O. Technology as a complex adaptive system: evidence from patent data. *Research Policy 30*, 7 (2001), 1019–1039.

10. Fowler, J.H. and Jeon, S. The authority of Supreme Court precedent. *Social Networks 30*, 1 (2008), 16–30.

11. Fowler, J.H., Johnson, T.R., Spriggs, J.F., Jeon, S., and Wahlbeck, P.J. Network Analysis and the Law: Measuring the Legal Importance of Precedents at the U.S. Supreme Court. *Political Analysis 15*, 3 (2007), 324–346.

12. Gomes-Casseres, B., Hagedoorn, J., and Jaffe, A.B. Do alliances promote knowledge flows? *Journal of Financial Economics 80*, 1 (2006), 5–33.

13. Hall, B.H., Jaffe, A., and Trajtenberg, M. Market value and patent citations. *RAND Journal of economics*, (2005), 16–38.

14. Harhoff, D., Narin, F., Scherer, F.M., and Vopel, K. Citation frequency and the value of patented inventions. *Review of Economics and statistics 81*, 3 (1999), 511–515.

15. Hart, H.C. Re: Citation System for Patent Office. *Journal of the Patent & Trademark Office Society 31*, (1949), 714.

16. Hicks, D., Breitzman, T., Olivastro, D., and Hamilton, K. The changing composition of innovative activity in the US—a portrait based on patent analysis. *Research policy 30*, 4 (2001), 681–703.

17. Jaffe, A.B., Trajtenberg, M., and Henderson, R. Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *The Quarterly Journal of Economics 108*, 3 (1993), 577–598.

18. Karki, M. Patent citation analysis: A policy analysis tool. *World Patent Information 19*, 4 (1997), 269–272.

19. Lemley, M.A. and Sampat, B. Examiner Characteristics and Patent Office Outcomes. *Review of Economics & Statistics 94*, 3 (2012), 817–827.

20. Lynch, M.J. Citators in the Early Twentieth Century-Not Just Shepard's. *Legal Reference Services Quarterly 16*, (1998), 5–22.

21. Marx, S.M. Citation Networks in the Law. *Jurimetrics Journal 10*, 4 (1970), pp. 121–137.

22. Narin, F., Rosen, M., and Olivastro, D. Patent Citation Analysis: New Validation Studies and Linkage Statistics. In A.F.J. Van Raan, A.J. Nederhof and H.F. Moed, eds., *Science Indicators: Their Use in Science Policy and Their Role in Science Studies*. DSWO Press, The Netherlands, 1988.

23. Office of Technology Assessment & Forecast, U.S. Patent & Trademark Office.*Technology Assessment & Forecast, Sixth Edition.* 1976.

24. Rosell, C. and Agrawal, A. Have university knowledge flows narrowed?: Evidence from patent data. *Research Policy 38*, 1 (2009), 1 – 13.

25. Sampat, B.N. When Do Applicants Search for Prior Art? *Journal of Law and Economics 53*, 2 (2010), pp. 399–416.

26. Seidel, A.H. Citation system for patent office. *Journal of the Patent Office Society 31*, 5 (1949), 554.

27. Shi, X., Adamic, L.A., Tseng, B.L., and Clarkson, G.S. The impact of boundary spanning scholarly publications and patents. *PloS one 4*, 8 (2009), e6547.

28. Smith, T.A. The Web of Law. *San Diego L. Rev. 44*, (2007), 309.

29. Thompson, P. Patent Citations and the Geography of Knowledge Spillovers: Evidence from inventor-and examiner-added ciatations. *Review of Economics & Statistics 88*, 2 (2006), 383–388.

30. Tijssen, R.J.W. Global and domestic utilization of industrial relevant science: patent citation analysis of science–technology interactions and knowledge flows. *Research Policy 30*, 1 (2001), 35 – 54.

31. Trajtenberg, M. A penny for your quotes: patent citations and the value of innovations. *The Rand Journal of Economics*, (1990), 172–187.

32. Uzzi, B., Mukherjee, S., Stringer, M., and Jones, B. Atypical Combinations and Scientific Impact. *Science 342*, 6157 (2013), 468–472.

33. Whalen, R. Bad Law Before it Goes Bad: Citation networks and the life cycle of overruled precedent. In R. Winkels, N. Lettieri and S. Faro, eds., *Network Analysis in Law*. ESI, 2013.

34. Wuchty, S., Jones, B.F., and Uzzi, B. The Increasing Dominance of Teams in Production of Knowledge. *Science 316*, 5827 (2007), 1036–1039.