Boundary Spanning Innovation and the Patent System: Interdisciplinary Challenges for a Specialized Examination System

Ryan Whalen

University of Hong Kong Faculty of Law

Abstract

This article discusses the importance of boundary spanning innovation, demonstrates the drawbacks of popular meta-data based boundary spanning measures, and proposes a new full text semantic similarity measure of boundary spanning. It subsequently uses the semantic distance boundary spanning measure to demonstrate that boundary spanning innovation has become more common in recent decades, and show that these boundary spanning inventions pose challenges for the traditional specialized-examiner patent examination model. Examining the applications for inventions that span technical boundaries takes longer and requires more back-and-forth with the patent office than their comparatively simple peers. Finally, this article discusses potential reforms to the patent examination system to help address these challenges.

Keywords: Patents; Intellectual Property; Citation Analysis; Boundary Spanning; Patent Examination; Innovation Policy

JEL Codes: O32, O33, O34, O38

Highlights:

- Defines new semantic-similarity citation measures of boundary spanning
- Empirically demonstrates boundary spanning inventions have increased over time
- Empirically demonstrates that patent examination has become more specialized recently
- Empirically demonstrates that boundary spanning patents take more time and more communication to assess
- Suggests patent assessment policy responses to changing innovation realities

1 Introduction

In recent decades, science and technology policymakers have promoted interdisciplinary research that spans the traditional boundaries between research communities (Cummings & Kiesler, 2005). Crossing disciplinary and technical boundaries has resulted in a variety of new areas of scientific development and new consumer goods. For instance, the increasingly important biotechnology sector draws on the traditionally distinct fields of biology and engineering. Likewise, consumer goods in recent years have integrated technologies from what were once distinct technical domains. While our in home temperature control was once accomplished with relatively simple mechanical switches, we now have thermostats that feature wireless connectivity, machine-learning capabilities, and remote control user interfaces.

Because of the potential that crossing disciplinary boundaries has for generating new high-impact ideas, policymakers remain interested in promoting research that brings together diverse sets of researchers. However, despite this increased focus on encouraging porousness in disciplinary boundaries, we have an incomplete understanding of the extent to which there has been more boundary spanning scientific output, and what effects this may have on innovation policy players such as patent offices. There is some evidence to suggest that boundary spanning has increased in recent years (Porter & Rafols, 2009), and that research spanning across disciplinary boundaries has outsized impact (Shi, Adamic, Tseng, & Clarkson, 2009). However, there has been little thought given to how this trend towards increasing boundary spanning and interdisciplinarity might affect how we incentivize and reward research behavior. Research suggests that interdisciplinary proposals have lower success in attaining funding support (Bromham, Dinnage, & Hua, 2016). This may occur at least partially as a result of it being more difficult to assess interdisciplinary proposals given that they do not fit neatly within expected knowledge frameworks. If a similar process holds in the assessment of technical information—that is, if it is more difficult for patent examiners to assess boundary spanning inventions—an increase in boundary spanning inventions may pose challenges for effective patent application assessment.

Effective patent examination is central to the modern innovation incentive system. Intellectual property law provides for limited monopolies on inventions, provided those inventions meet the patentability threshold. There is evidence to suggest that patent offices are already straining under the application workload, which has increased dramatically in recent years. Frakes and Wasserman (2015) show that as patent examiner workload increases, so too does their propensity to grant bad patents. This research focuses on the increasing number of patent applications, and the effect this can have on 21st century patent offices. However, it provides little insight into whether patent applications have changed qualitatively as well as quantitatively, and if so, what effects this might have on our ability to effectively examine 21st century patent applications.

Due to the increasing popularity of interdisciplinary research, one primary way that inventions may have qualitatively changed in recent decades, is in an increasing tendency to span disciplinary boundaries. This could arise as the product of positive campaigns to encourage interdisciplinarity (Haythornthwaite, 2006), or due to fortuitous boundary spanning discoveries, as a response to market demand for more boundary spanning products, or as a result of information technology that facilitates the discovery of ideas that researchers might not otherwise encounter (Whalen, 2015). To determine whether inventions have changed qualitatively in recent decades, and what if any implications this has on the patent system, we first need to discuss the relationship between boundary spanning and innovation more generally.

2 Boundary Spanning & Innovation

Spanning boundaries has long been associated with good ideas (Burt, 2004), and high impact scientific and technical developments (Leahey, Beckman, & Stanko, 2015; Shi et al., 2009; Tushman & Scanlan, 1981). Boundary spanning occurs when researchers draw on expertise from distinct and disparate fields. This enables them to engage in a sort of knowledge arbitrage as they broker information across boundaries (Tushman, 1977; Tushman & Scanlan, 1981). By doing so, researchers are more likely to discover novel connections, and generate ideas or inventions offering unique solutions or capabilities.

Making these combinations across rarely combined fields leads to a higher probability of a scientific article or patent becoming high impact (Uzzi, Mukherjee, Stringer, & Jones, 2013). Similarly, using a diverse interdisciplinary mixture of knowledge inputs is also associated with higher impact (Chen, Arsenault, & Larivière, 2015). However, spanning technological boundaries is not without cost. Doing so increases success variance, leading both to more low and high impact outcomes (Fleming, 2001; Yegros-Yegros, Rafols, & D'Este, 2015).

At least in the context of academic science, spanning disciplinary boundaries has increased in recent decades (Porter & Rafols, 2009). Much of this has occurred as our collective knowledge has become so vast that it has become increasingly important for researchers to team together with one another in order to assemble sufficient knowledge mastery (Jones, 2009). Meanwhile, developments in information and communication technologies have enabled both increased teamwork and eased research, facilitating innovation (Whalen, 2015). As this has occurred, both the frequency and the impact of team research have increased (Wuchty, Jones, & Uzzi, 2007), and those teams are now more likely than ever to cross disciplinary boundaries (Porter & Rafols, 2009).

As research that spans across disciplines has increased, there has been a wide variety of discussions about the implications this has on various facets of science and research. Scholars have explored the implications on universities (e.g. Lattuca, 2001), academic publishing (e.g. Rafols, Leydesdorff, O'Hare, Nightingale, & Stirling, 2012), and applied research (e.g. Etzkowitz, 1998). There has been comparatively little attention paid to patent offices, and whether an increase in boundary spanning inventions has occurred and if so, what implications, if any, that may have for the way we incentivize inventions and assess patent eligibility. This leads me to pose the following research question:

RQ1: Has there been a change in the tendency for patents to span boundaries, bringing together distant knowledge?

2.1 Boundary Spanning and the Patent System

Patent examination is the primary task of patent offices. It is a labor-intensive task, requiring patent offices to hire large corps of examiners, train them extensively, provide them with advanced information resources. It is ultimately a costly endeavor, with the USPTO requiring an annual budget of approximately $3.3 billion, while the EPO has a budget of approximately €2 billion. Increasing boundary spanning innovation has the potential to complicate the traditional specialized examiner model used by most patent offices, and thereby potentially upset the existing examination regime.

2.1.1 Patent Examination Specialization

In many jurisdictions, the patent application examination process has long been, and largely remains, an individual one. Examiners are assigned applications, and depending on their seniority and the rules of their patent office, often do the majority of their patentability assessment with little input. Their work is at times reviewed by a supervising examiner, but for the most part the process proceeds individually. In the United States, this has largely been the case since the Patent Office was founded in 1836. Early examiners were generalists, and expected to be able to assess applications in any technological area (Post, 1976). Over time, as technology grew more intricate and complex, Patent Offices encouraged specialization, establishing technology centers and art units with expertise in specific technical areas.

We see this increasing specialization empirically when we look to the examination loads of particular patent examiners. The USPTO Patent Application Information Retrieval (PAIR) data includes unique examiner IDs and records of the technological classification of the applications they have assessed. Using this, we can associate each examiner with their assigned workload and subsequently calculate the average breadth of an examiner's work over time. By breadth, I refer to the number of distinct USPC subclasses that each individual examines at least one application within. Calculating this for every active examiner on a yearly basis, shows us how many subclasses the average patent examiner worked within each year (see Figure 1).

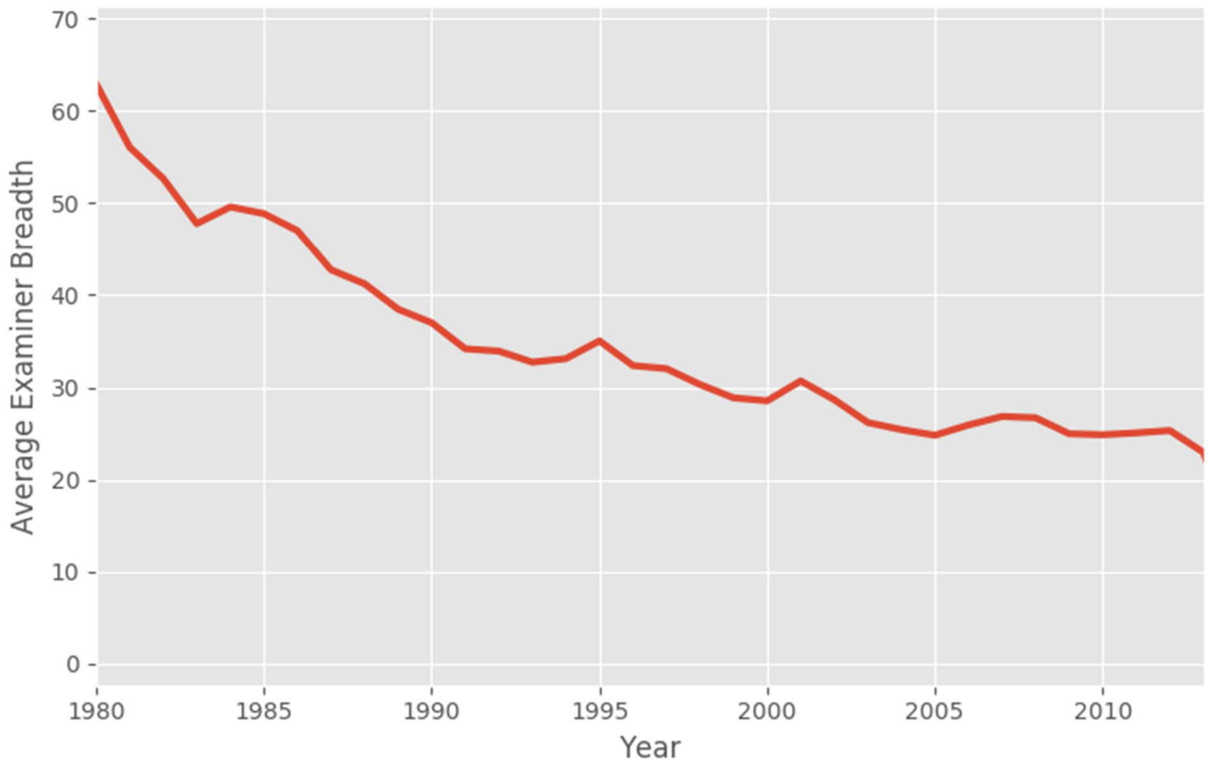


Figure 1. The average number of subclasses examined by a patent examiner over the course of a year's work. This includes data on 15,333 unique patent examiners' work on 7,842,980 applications.

We see in Figure 1 that in recent decades there has been a steady decrease in the number of patent classes that each examiner works within. This suggests that examiners are becoming more narrowly-focused on specific areas of technical expertise. For the most part, this increase in specialization has been a good thing, allowing for efficient examination of highly-specialized technologies. Specialization allows for more familiarity with the relevant prior art and fluency with the associated technical language. However, specialization's strengths can become a weakness if inventions do not fit clearly within a pre-defined technological area, but instead draw on inspiration from diverse fields.

Boundary spanning inventions by definition do not fit neatly within disciplinary boundaries. Instead, they draw on diverse sets of information and span multiple technical areas. We see in scientific research funding applications that interdisciplinary research proposals are less-likely to receive funding than proposals that fit more neatly within traditional disciplinary boundaries (Bromham et al., 2016). Although this could arise if interdisciplinary research is of consistently lower quality, it could also be the product of inherent challenges that interdisciplinary research poses for traditional methods of assessing research merit (Feller, 2006). In the case of boundary spanning research, the specialized backgrounds and training of those who assess the research may become a liability rather than an advantage. It is reasonable to expect that this same principle may be at play in the patent examination context. Given that examiners are increasingly domain experts with specialized knowledge, they may be at a disadvantage when assessing inventions that span technical boundaries.

Both interviews with patent examiners and quantitative assessment of pendency times suggest that particularly complex inventions generally require more time to assess (Popp, Juhl, & Johnson, 2004). These “complexity problems are particularly acute in cross-disciplinary fields” because of the demands that cross-disciplinary inventions pose for patent examiners as they attempt to search the prior art and assess patentability (Popp et al., 2004, p. 9). Although it is probably true that not all boundary spanning inventions will pose complexity challenges for patent examiners, because of the way they re-combine information from diverse sources they are more likely to do so. By bringing together distantly-related information inputs, boundary spanning inventions require examiners to search more areas of prior art, and compare the application’s claims against existing knowledge in more than a single field. This leads me to hypothesize that:

H1: Patent applications that span boundaries will have longer pendency periods

Similarly, these more complex boundary spanning inventions may require examiners to communicate more often with applicants, whether it be with requests to clarify details about the invention, amend claims, or other reasons associated with patent prosecution. The hypothetical mechanism driving this effect is largely the same as that hypothesized above: inventions drawing on distantly related information antecedents ask more of examiners and in doing so increase the chances that those examiners need to communicate with applicants, whether it be to ask for clarification of an unclear claim, suggest dividing the application into multiple inventions, or otherwise. Indeed, existing research suggests that technical fields drawing on multiple areas of knowledge often require "more frequent communications" between applicants and examiners (Popp et al., 2004, p. 41). This leads to the second hypothesis:

H2: Patent applications that span boundaries will require more transactions between the patent office and the applicant prior to approval.

Testing these hypotheses requires a measure to detect boundary spanning. However, despite the importance of boundary spanning and its increase in recent years, there are few agreed-upon or established methods to detect boundary spanning inventions and their associated patents, and those that do exist share a variety of weaknesses. The next section will discuss the methods most commonly used, and their weaknesses, before introducing a method that uses the full text of granted patents to detect those that draw on particularly diverse inspiration.

2.2 Measuring Boundary Spanning

Boundary spanning is fundamentally a network concept. It shares traits with the concepts of structural holes (Burt, 2004) or weak ties (Granovetter, 1973). These network concepts each provide a somewhat different perspective on network bottlenecks and their importance in the flow of information. As a network concept, boundary spanning can be operationalized on a variety of types of networks, including social networks and information networks. In social network terms boundary spanning occurs when individuals link together organizations or individuals and translate information across the boundaries that separates them (Tushman & Scanlan, 1981). In information network terms, boundary spanning occurs when documents link together disparate or dissimilar pieces of information (Shi et al., 2009).

As this project's primary focus is identifying and measuring boundary spanning inventions, I will focus on an information network perspective on boundary spanning. Existing approaches to measuring boundary spanning within information networks largely rely on document metadata—especially categorization and citation relationships—oftentimes in conjunction with one another. These metadata have been used in a variety of different ways to detect when documents span boundaries within the information network.

One approach examines the combination of reference types referred to within the focal document. This "atypicality" approach infers the content referred to based on the category of either the journal (Uzzi et al., 2013) or patent technology class (Fleming, 2001) cited. Documents that draw on atypical combinations can be considered boundary-spanners of sorts. The inference here is that atypical combinations bring together bodies of knowledge that are rarely combined, spanning their boundaries by bringing together rarely combined types of content. Shi and colleagues use a related method in their work exploring the impact of boundary spanning journal articles and patents (Shi et al., 2009). However, rather than focusing on the combination of categories referenced, they look to the likelihood of observing a citation relationship between the category of the citing document and that of the cited document.

The "diversity" approach is in some ways similar to the atypicality approach but conceptually distinct. Rather than measuring how typical or atypical a given mixture or pair of referenced journals or technology classes is, the diversity approach looks to the mixture of different types of journals or technology classes cited and from those calculates a diversity index

(Bromham et al., 2016; Chen et al., 2015). The inference here is that documents with particularly diverse sets of information inputs, span disciplinary boundaries.

Using categorization data itself provides another way to infer boundary spanning. Multiple categorizations for a document, especially when those categories rarely appear together can suggest that it spans knowledge boundaries. Mapping co-categorization frequency can provide insight into both the underlying structure of a knowledge space as well as the extent to which a given document makes novel connections within that. This can be used to better understand a knowledge domain (Tijssen, 1992), detect convergence between previously distinct knowledge areas (Curran & Leker, 2011), or explore the knowledge generation process (Youn, Strumsky, Bettencourt, & Lobo, 2015).

These metadata approaches to detecting boundary spanning have a number of strengths. They are relatively simple to calculate, and easy to understand. However, they also have a number of weaknesses. Because typicality constantly changes, based on the continuous submission of new patent applications, models require constant updating. These approaches also require consistent classification systems applied rigorously. In the context of patents, differences in the way categories are assigned, especially differences in how likely they are to be assigned multiple secondary classifications, can be problematic for these metadata approaches. Co-cateogirzation approaches, which leverage these multiple categorizations, are themselves subject to a number of shortcomings, perhaps most importantly the fact that they are undefined in instances when a document receives only a single classification—the case in approximately 19% of the data analyzed below.

In the context of journals, accurately categorizing some of the most influential general interest scientific journals such as *Science* and *Nature* is challenging. Perhaps the greatest challenge for metadata approaches to boundary spanning detection lies in the nature of metadata. It is by definition coarse and somewhat abstract. These approaches lump publications within the same category together, essentially glossing over all of the intra-category variety that exists. Furthermore, when documents are given multiple categorizations as often occurs for patents, most of the existing research glosses over this nuance and simply assumes the primary classification adequately represents the document's contents.

These metadata-based boundary spanning measures are emblematic of the "metaknowledge" revolution that has occurred in recent years (Evans & Foster, 2011). The metaknowledge revolution can be attributed to increased capabilities in dealing with large datasets, and improved access to databases featuring publication metadata. Although the metaknowledge revolution has made significant contributions to our understanding of the research and knowledge production processes, its focus on metadata is perhaps too narrow in an era when access to more detailed data is available. In recent years, there has been extensive growth in techniques for analyzing natural text as well as access to larger-and-larger sets of full text documents. Research suggests that semantic approaches using document content can outperform metadata based approaches (Preschitschek, Niemann, Leker, & Moehrle, 2013). However, there have been few attempts to use this access to text and these natural language processing techniques to measure the extent to which publications span boundaries. Using the text of the documents to assess their contents rather relying on the metadata categories they are assigned as a proxy for content avoids many of the challenges that metadata approaches face, and leads to more potential nuance in boundary spanning measures. I thus propose a text-based measure to identify boundary spanning patents in the following section.

3 Data & Methods

A full text approach to detecting document boundary spanning will ideally replicate many of the strengths of the existing metadata approaches, while also addressing some of their weaknesses. Whether they measure atypicality of references, the likelihood of observing citations from one category to another, co-categorization, or the diversity of categories referenced, each of the metadata approaches seeks to identify documents that unite distant or diverse knowledge areas. These approaches, implicitly or explicitly, adopt a spatial understanding of knowledge, taking the position that some pieces of knowledge are closely related, while others are only distantly so. Spatial knowledge metaphors are common, with scholars referring to proximate knowledge search as “local” (Stuart & Podolny, 1996), “intensive” (Jovanovic & Rob, 1990), or “exploitative” (March, 1991), while more distant search has been referred to as global, “extensive” (Jovanovic & Rob, 1990), or “explorative” (March, 1991). Thus, a full text approach to detecting boundary spanning must be able to measure the “distance” between documents. This happens to be one of the strengths of a full text approach in that there are a variety of established measures of semantic distance or similarity that can be leveraged to provide nuanced measures of how distant documents are from one another. Using the full text of documents allows one to actually compare their content to determine how similar they are rather than relying on the necessarily coarse proxy measures provided by metadata.

To leverage the power of full text to help us measure the boundary spanning, I draw on the full text and citation data of all utility patents granted by the USPTO from 1976 until late 2014. Using the text from the abstract, claims, and description of each of these patents, I then compute a 500 dimension latent semantic analysis model (Deerwester, Dumais, Landauer, Furnas, & Harshman, 1990; Landauer, Foltz, & Laham, 1998).[1] Latent semantic analysis, enables document comparison and allows for the detection of latent similarities between them. Unlike simpler document vector space methods, LSA does not require documents to contain exactly the same words to detect similarities. If words are used in similar contexts (e.g. car and automobile) LSA is able to detect these contextual similarities and treat the words similarly. We can consider the 500 dimensions that results from the LSA model as concepts, with the corresponding scores for each document representing the degree to which that document is related to that concept.

Once the LSA model is computed, each patent can then be located with the resulting 500-dimensional space, and cosine distance can be used to measure how distant they are from one another. With the LSA coordinates established, boundary spanning patents can be detected by examining cited prior art references. Patent prior art references have been used by scholars in a variety of ways. Citations have been used as measures of an invention's value (Trajtenberg, 1990), and have been shown to correspond to the patenting firm's market value (Hall, Jaffe, & Trajtenberg, 2005), and objective measures of the effectiveness of the patented technology (Moser, Ohmstedt, & Rhode, 2015). Citations have previously been used to detect boundary spanning, but only via methods that rely using on metadata categorizations to infer content and subsequently measuring the likelihood of observing a citation from one category to another (Shi et al., 2009).

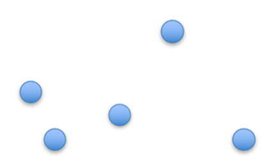
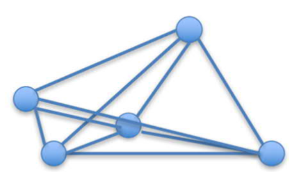
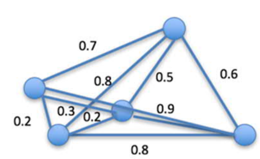
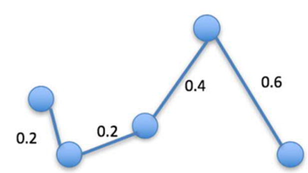
[1] To prepare the data for the LSA, extremely common (i.e. occurring in more than 50% of the documents) words are removed from each document because they provide little insight into content. Similarly, extremely rare words (i.e. occurring in fewer than 5 total documents) are also removed, because they are likely to be typographical errors. A tf-idf transformation is then used on the corpus to re-weight words from raw counts to tf-idf values. These tf-idf values are then used to compute the LSA model.

By focusing on prior art references in the full text boundary spanning detection method, we can gain insight into the body of knowledge that is related to a new invention, providing perspective on the expanse of the knowledge space that it spans. Patent prior art often provides a source of knowledge for inventors to build upon (Sternitzke, 2009). The citation relationships that exist, although not always added by inventors (Alcácer, Gittelman, & Sampat, 2009), can reveal relationships between technologies and suggest flows of knowledge (Almeida & Kogut, 1999, 1999; Jaffe, Trajtenberg, & Henderson, 1993; Rosell & Agrawal, 2009). These cited prior art inventions thus provide a useful starting point to determine the degree to which a new invention spans technological boundaries.

To measure boundary spanning, I focus on each patent's prior art references and the distances between them. If a patent cites to prior art that is universally similar to one another—for example, a patent for an improved travel coffee cup citing to five previous coffee cup patents—the distance between the prior art references will be relatively low, suggesting that the invention does not span technical boundaries. If on the other hand, an invention draws on art from diverse technical areas—for example an improved coffee cup patent that cites to both coffee cup prior art and hot water heater prior art—the distances between the prior art references will be greater.

To measure boundary spanning, the resulting co-cited prior art distances are used to assemble a co-cited network. In this network, each node is a cited prior art reference, and the links between them are weighted by their pairwise semantic distance. By definition, this co-cited network is completely connected. To get a sense of the total distance within semantic space covered by a patent's prior art co-citation network, we can take the minimum spanning tree-representing the shortest path that connects each of the cited documents. This minimum spanning tree shows the shortest path that will traverse all of the cited documents within the co-cited subgraph. As such, if we consider it from the technical knowledge space perspective, it shows the shortest distance separating the different technical knowledge areas that have been referenced by the focal patent. The maximum distance in this network represents the greatest leap between technical areas, and thus provides insight into the degree to which an invention spans disparate technical areas.

Table 1. Showing the process for measuring boundary spanning based on the semantic distance between co-cited prior art references.

Step 1: Identify all co-cited prior art patents	Step 2: Create a fully-connected network of all co-cited patents	Step 3: Weight the network based on semantic similarity	Step 4: Find the minimum spanning tree
			

This method results in high boundary spanning scores for inventions which cite to disparate technological fields. For instance, patent number 6,253,505 (claiming an "Iron-oxide-containing one or two-component polyurethane paint for coating elastomers, its production and use") spans a number of boundaries as demonstrated by citations to screw manufacturing technologies, glass composition technologies, and anti-friction paint technologies. As such, this boundary spanning patent scores quite highly in the co-citation distance measure. On the other hand, patent number 8,653,339 (claiming "Soybean variety XBP42005") cites only within the soybean technology field (including references to varieties XB42H07, XB31H07, and XB40K07) and thus scores very low on the co-citation distance measure.

4 Results

Figure 2 shows the distance distribution between randomly selected patent pairs demonstrating that the baseline distance between any two documents in the corpus is relatively high. Given the wide variety of technical areas covered by patents, this is unsurprising. We would not expect any two randomly selected patents to have much in common. On the other hand, Figure 3 shows the distance distribution for co-cited prior art references. This tells a very different story, revealing that prior art that is cited by the same patent tends to be much more similar than the baseline similarity we see in the entire patent corpus.

Figure 2. Distribution of pairwise patent distances

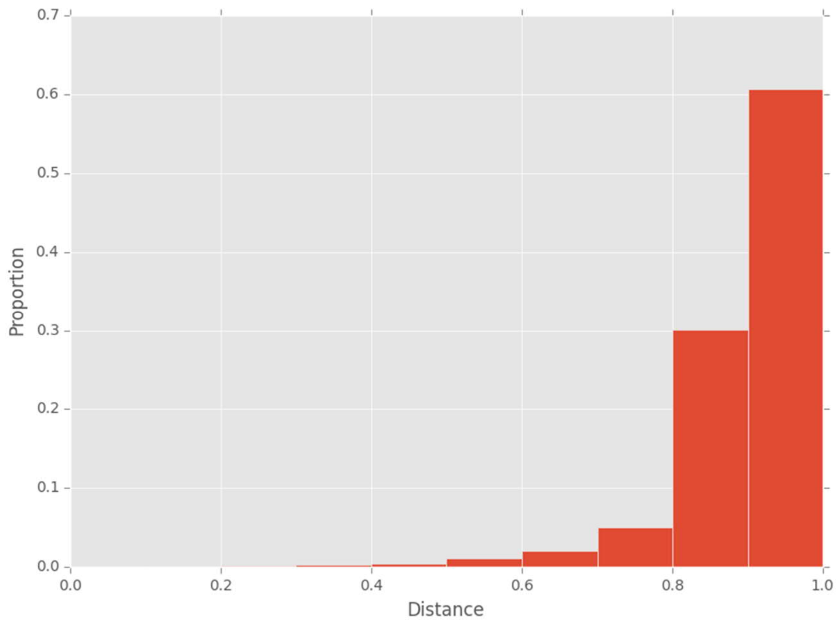
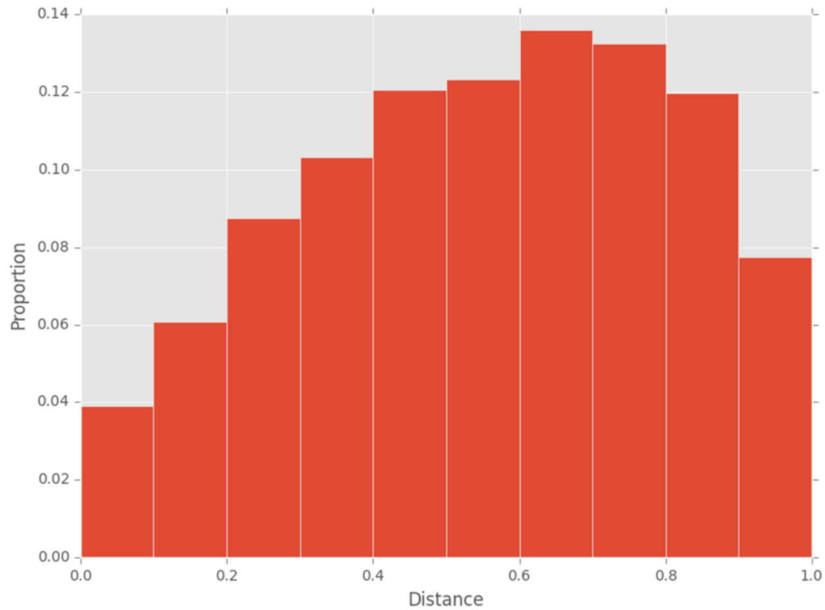
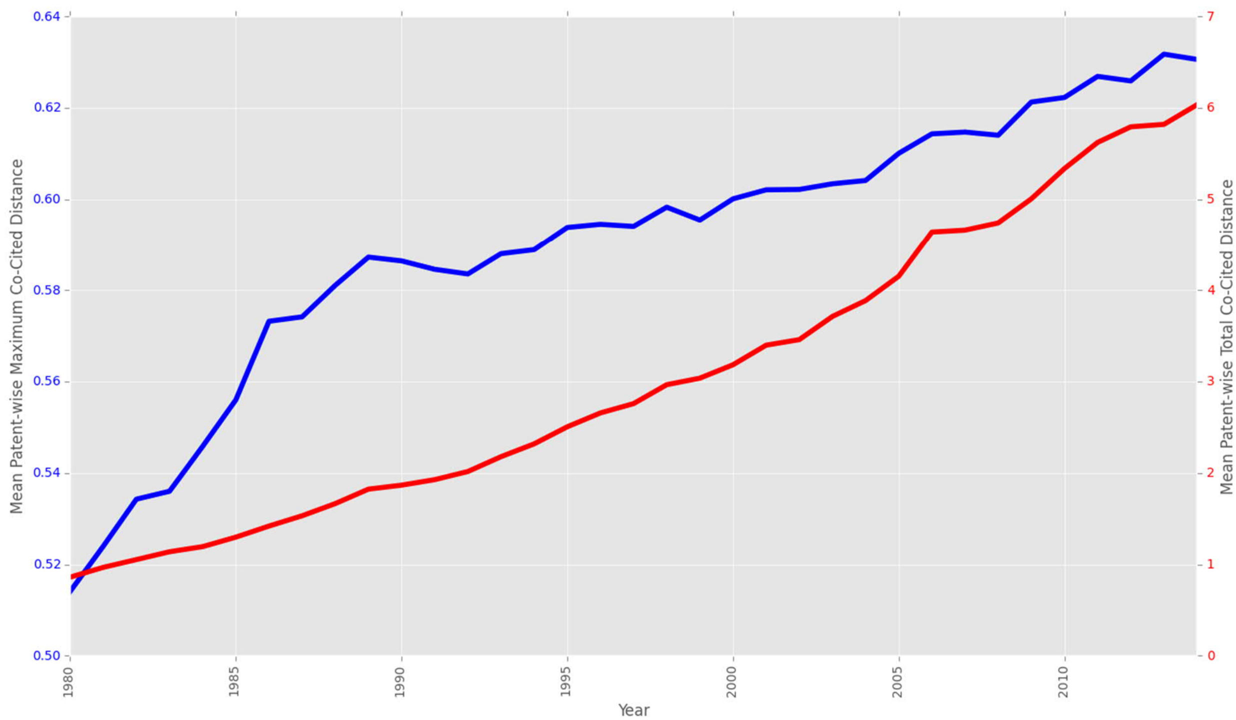


Figure 3. Distance distribution of co-cited patent pairs.



Examining the boundary spanning trend over time (Figure 4) shows consistent increase in both the maximum distance between co-cited documents, as well as the total semantic distance covered by the co-cited prior art network. This suggests that recent patents are more likely to cite semantically distant documents, bringing together diverse information inputs as they span technical boundaries. Although the increase in the total semantic distance observed in the co-cited network (the red line) can be partially explained by an increase in the average number of prior art citations included in patents, the increasing maximum co-cited distance demonstrates that patents not simply citing more prior art, they are citing more diverse prior art. These results provide an answer to RQ1, showing that there has been a steady trend towards patents bringing together more-and-more distant knowledge, suggesting that inventors are now reaching “further” away in the knowledge space for antecedent knowledge and spanning more distant knowledge areas.

Figure 4. The maximum co-cited semantic distance (blue) and mean total co-cited semantic distance over time.



One might be concerned whether the above increase in boundary spanning is the result of changes in patent drafting strategies rather than an actual increase in boundary spanning inventions. To provide further insight into changes in inventions over time, we can take an alternate approach to assessing how many fields inventions draw upon. Although they are subject to some weaknesses, especially when assessing inventions at the single patent-level, a metadata approach here can us better understand whether or not the increase in boundary spanning that the semantic approach above suggests has occurred has actually done so. Figure 5 shows the average number of distinct technology classes cited by patents over time. This demonstrates that in recent decades inventions have tended to cite to more-and-more technological areas, providing further support for the observation that the incidence of boundary spanning inventions has indeed increased.

Figure 5. Average number of unique CPC subclasses cited by patents granted in each year.

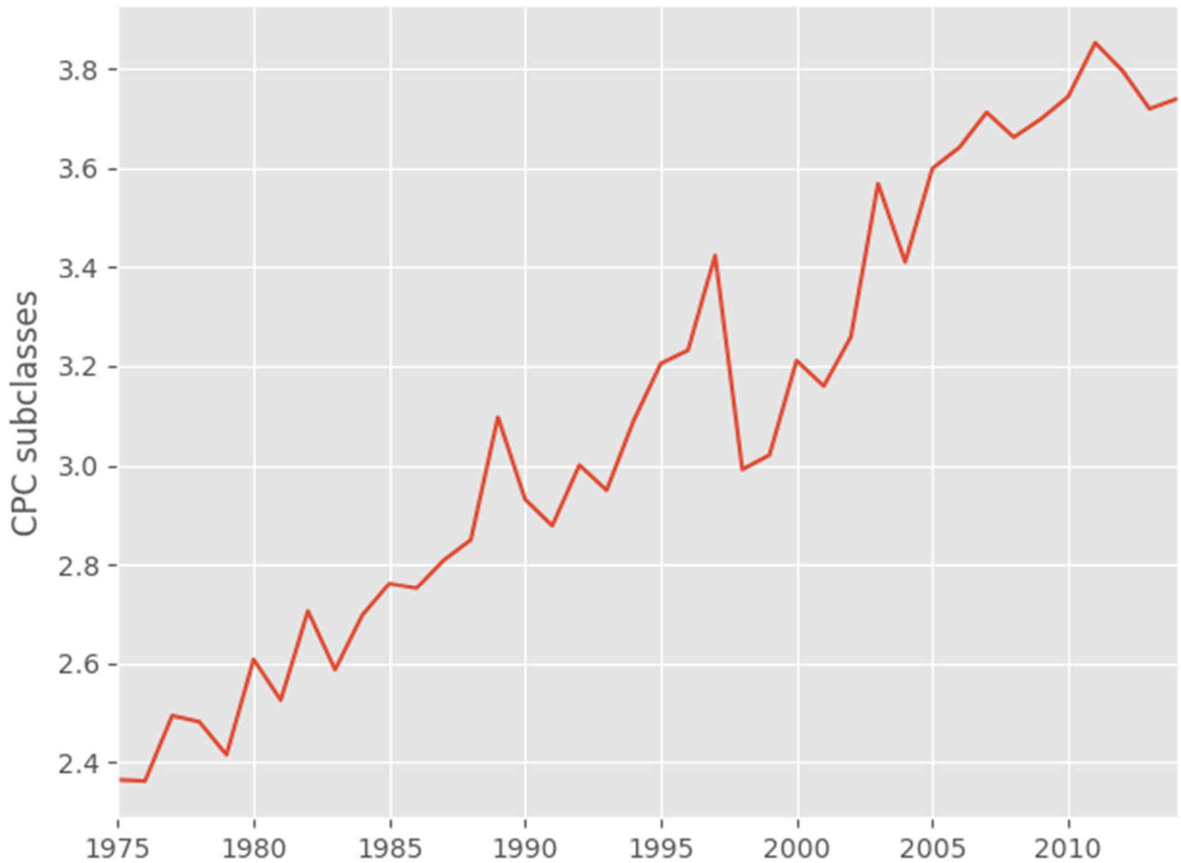


Table 2 shows the descriptive statistics for both the outcome, predictor and numeric control variables used in the below regression models. Checking for collinearity among the variables shows no pairwise correlation exceeds $p = 0.18$. Table 3 shows the results of OLS regression models[2] examining the relationship between boundary spanning—here defined as the maximum distance in the co-cited network as described above—and both the time a patent application is pending for prior to grant, and the number of transactions between the applicant and the patent office.

[2] Robustness checks using cox-hazard, poisson and negative binomial models show substantively similar results. Further robustness checks using an operationalization of boundary spanning that counts the unique number of CPC subclasses listed on patents cited by the focal patent also show substantively similar results.

Table 2. Descriptive Statistics

	Mean	Median	Standard Deviation
Pendency	925.5	803.0	513.35
Transactions	34.55	31.0	20.85
Maximum co-cited distance	0.6	0.6	0.21
Independent claims	15.47	13.0	12.39
Patent references	12.72	7.0	32.54

Table 3

	Dependent variable:	
	Pendency	Transactions
	(1)	(2)
Maximum co-cited distance	88.087***	4.184***
	(1.189)	(0.038)
Number of independent claims	2.868***	0.110***
	(0.019)	(0.001)
Number of patent references	0.558***	0.089***
	(0.007)	(0.0002)
Year control	Yes	Yes
Category control[3]	Yes	Yes
Constant	512.282**	-0.603
	(254.165)	(8.217)
Observations	3,623,126	3,623,126
Adjusted R^2	0.278	0.539
Note:		*p**p***p<0.01

These results provide support for both H1 and H2. We see that as boundary spanning increases so too, does the number of days of pendency for the related patent application. Similarly, as boundary spanning increases the patent office has an increasing number of back-and-forth transactions with the applicant. This suggests that these boundary spanning inventions present an increased workload for patent examiners, and consequently has implications for the

[3] This category control, uses an updated version of the 37 super-categories proposed in the NBER patent dataset (Hall, Jaffe, & Trajtenberg, 2001).

patent system. These results hold when controlling for other aspects that one would expect to predict both pendency and the number of transactions, including: the number of claims made in the patent, the total number of prior art references, the type of technology claimed, and the year the patent application was granted.

5 Discussion

The above results leverage the full text of patents and their prior art citation records to demonstrate that the tendency for inventions to integrate distant knowledge has increased in recent decades. Furthermore, these boundary spanning inventions take longer for the patent office to examine, and require more back-and-forth between applicants and examiners. This has a variety of implications for the patent examination system.

5.1 Implications for patent examination

One of the challenges facing effective assessment of boundary spanning patents arises from the way patent examiner work is credited by patent offices. In the U.S. Patent Office, the working time that examiners are credited with when examining an application is a function of the technology classification of the invention, the patent examiner's seniority, and the type of office action taken (Popp et al., 2004). By taking into account the classification of the invention, the count system attempts to credit examiners' work according to the average complexity of inventions within that category, with more complex technologies being granted more time for the examination process. However, there is much heterogeneity within technology classes. For example, compare two patent applications in class 725[4]: application numbers 10,100,643 and 10,434,042. The first of these (the '643 application) claimed a "Multimedia display system using display unit of portable computer, and signal receiver for television, radio, and wireless

[4] United States Patent Class 725 denotes Interactive Video Distribution System inventions

telephone." The second (the '042 application) claimed a "Method and apparatus for browsing using multiple coordinated device sets." Because both of these applications were in technology class 725 (interactive video distribution systems) they would each have an expectancy of 31.6 hours.[5] However, they are starkly different in their levels of complexity and the amount of work they would have required to process. For instance, the '643 application had a description that was 3,443 words long and included 2 independent claims and 3 dependent claims. The specification for this application ran 10 double-spaced pages. Meanwhile, the '042 application's description ran 110,483 words long and included 37 independent claims and 247 dependent claims. This longer application's specification was 247 pages long as submitted.

This intra-class heterogeneity is exacerbated when an invention spans technological boundaries. Although the USPTO will, and indeed often does, assign multiple categories to an application, each application and granted patent is given a single primary classification. However, when an invention draws on disparate knowledge and disparate prior art, the examination task is that much more complex, requiring more time for a thorough prior art search and patentability assessment. The current count regime is inflexible to these demands. If patent offices were able to detect boundary spanning inventions, they would be able to better ensure that these particularly challenging—and potentially important—applications are given the time they are due.

This challenge of appropriately compensating examiners for the work they do could be partially addressed by adopting a "complexity based count system." Such a system would take into account a variety of factors relevant to an application's complexity—e.g., application length, number of claims, USPC, citation distance, etc.—and use them to vary the number of hours an

[5] This is the base expectancy for class 725. The actual expectancy is calculated by dividing the expectancy by the position factor, which is a function of a patent examiner's rank.

examiner is credited with for disposing of the application. Officials at the USPTO appear to be considering such a move. In a recent report on efforts to ensure patent quality, directors of two technology centers indicated that increasing technological complexity posed a challenge for examiners, and that re-structuring examination time allowances may offer one way of addressing the challenge (Sullivan & Lefkowitz, 2017).

An increase in boundary spanning inventions also poses challenges for the Patent Office's organization into specialized subunits. The USPTO's examination unit is currently structured into nine specialized technology centers, each of which is comprised of dozens to hundreds of art units. This art unit structure is a product of the trend towards increased specialization that the Patent Office has adopted since it was founded in the early 19th century. For the most part, this specialized organizational structure is a strength, allowing for efficient assessment of applications that fall within an art unit's area of expertise. However, as new combinations of secondary classifications grow (Youn et al., 2015), and inventions draw on increasingly distant and disparate antecedent information, a specialized approach to examination becomes more-and-more likely to be a liability. To address this challenge to the its traditional specialization-oriented structure, the Patent Office should consider opening avenues for application examination across art units. If an invention spans boundaries and substantially implicates technologies in multiple fields, the patentability assessment may differ across examiners working in different fields. By allowing for patentability input from multiple art units the Patent Office could help minimize the chance of granting bad patents due to insufficiently broad examination. This would mirror at the Patent Office the trend towards teamwork that we have seen in researcher generally (Jones, 2009; Wuchty et al., 2007). As the knowledge space has become ever-larger, researchers have been forced to work together in order to successfully navigate it. To date, the Patent Office has done little to increase its avenues for teamwork, and this may be a weakness given the norms of 21st century innovation.

The above proposed flexibility in examination hour crediting or patent examining unit organization would not come without cost and should thus be carefully considered. Adding flexibility that tailors examiner incentives or allows the patent office to more efficiently cross the institutional silos between art units runs the risk of increasing the workload for already overburdened patent examiners, adding to the costs incurred by the patent office, and prolonging the examination process. Instituting such substantial changes within an organization as large and economically important as the USPTO should not be done without substantial consideration, and ideally proof-of-concept testing. Before implementing these ideas pilot projects should be used to better understand both their effectiveness, and the challenges they might raise.

5.2 Methodological implications

In addition to its implications for patent assessment policy, this research also has methodological implications. As discussed above, increased access to large sets of metadata has led to a dramatic rise in "metaknowledge" research, exploring how knowledge is created and disseminated. In recent years, access to full text of research documents has similarly increased, with growth in the number of scientific articles published on open access archives, and public machine-readable patent datasets. This increased access to research documents, combined with reduced costs for data storage and computational capacity, set the stage for the growth of "computational knowledge research" marrying traditional knowledge research disciplines with techniques from computer and information science.

The method presented above, using the full text of patents to assess their semantic distance from one another and subsequently using that distance to inform citation analysis, shows promise in improving the way scholars detect boundary spanning, and also in the way other forms of citation analysis are employed. Citation analyses are used in a variety of research contexts. From backward citation studies that examine knowledge flows and recombination, to forward citation studies that assess impact and knowledge diffusion, citations have demonstrated a wide range of utility for researchers in a variety of fields. Adding semantic citation distance weighting to citation analyses can provide researchers with a more nuanced set of measures, improving their insight into the research and knowledge diffusion processes.

5.3 Limitations and Future Work

The full text semantic citation distance based approach to measuring boundary spanning used above has a number of strengths. It avoids the necessary coarseness associated with metadata-based measures. It allows for more nuanced measures, and intra-class variation. However, it does have some weaknesses. By relying on co-cited distances, the measure used above provides insight into the total area of the knowledge space implicated by a new invention. However, this—like all citation based measures—relies on prior art citations being included in patents. Why citations are included and why they are omitted is not always clear. Thus, citation-based metrics are sensitive to variations in the way prior art references are included or excluded from patent grants. Perhaps more importantly, relying on co-cited distance requires patents to have at least two prior art references in order to define a boundary spanning measure as proposed above. The majority of patents meet this threshold, but for those that do not, co-cited distance remains undefined.

More work is needed to better understand how boundary spanning inventions may pose challenges for the specialized examiner protocols used in most jurisdictions. The results reported above rely on granted patents, working backwards to both determine which span information boundaries, and how that relates to the length of time it takes for the Patent Office to grant them and how much back-and-forth they require in the process. Further work examining application data and delving in more detail into the nature of the transactions between applicants and examiners could help shed further light on the challenges that a trend towards interdisciplinarity may have for a specialized patent examination system.

Another area for future research is in further exploring semantic citation distance measures. The measure here demonstrates the utility of examining the semantic distance between co-cited documents. Examining other types of document relationships can provide different perspectives on the research and innovation process. Some work on backward citation distance demonstrates that scientists have tended to cite increasingly distant documents, especially following the introduction of the Internet and improved scientific search engines (Whalen et al., 2016). In addition to backward citation distance, forward citation distances and co-*citing* measures can provide insight into knowledge diffusion and perhaps more nuanced research impact measures.

6 Conclusion

In recent decades, patented inventions have tended to cite to more-and-more distant areas of the knowledge space. By spanning across distant knowledge domains, these inventions bring together ideas, technologies, and techniques from multiple areas, re-combining them to benefit by mixing this distant knowledge together. Although this increase in boundary spanning inventions can in some ways be considered a boon for contemporary consumers, it raises potential challenges for a patent examination system that has for decades focused on specialization.

The innovation policy incentives established by the patent system rely on effective and accurate patentability assessment. If patents are granted that should not be, these bad patents risk unnecessarily enclosing intellectual property, making follow on innovation more difficult and costly. Similarly, if deserving patents are not granted, the incentive to invest in innovation is undermined. It is thus essential that patent offices be responsive to changes in innovation norms. This has occurred throughout the history of the United States Patent Office, with its growth from two general examiners in 1836 to the large corps or specialized examiners that we see today. This responsiveness must continue into the 21st century to ensure that the examination practices of the patent office are adequate for assessing the increasingly boundary spanning innovation that we see today.

References

Alcácer, J., Gittelman, M., & Sampat, B. (2009). Applicant and examiner citations in U.S. patents: An overview and analysis. *Research Policy*, *38*(2), 415–427. doi:10.1016/j.respol.2008.12.001

Almeida, P., & Kogut, B. (1999). Localization of Knowledge and the Mobility of Engineers in Regional Networks. *Management Science*, *45*(7), 905–917. doi:10.1287/mnsc.45.7.905

Bromham, L., Dinnage, R., & Hua, X. (2016). Interdisciplinary research has consistently lower funding success. *Nature*, *534*(7609), 684–687. doi:10.1038/nature18315

Burt, R. S. (2004). Structural Holes and Good Ideas. *American journal of sociology*, *110*(2), 349–399.

Chen, S., Arsenault, C., & Larivière, V. (2015). Are top-cited papers more interdisciplinary? *Journal of Informetrics*, *9*(4), 1034–1046. doi:10.1016/j.joi.2015.09.003

Cummings, J. N., & Kiesler, S. (2005). Collaborative Research across Disciplinary and Organizational Boundaries. *Social Studies of Science*, *35*(5), 703–722.

Curran, C.-S., & Leker, J. (2011). Patent indicators for monitoring convergence - examples from NFF and ICT. *Technological forecasting and social change*. Retrieved from http://agris.fao.org/agris-search/search.do?recordID=US201301927988

Deerwester, S. C., Dumais, S. T., Landauer, T. K., Furnas, G. W., & Harshman, R. A. (1990). Indexing by latent semantic analysis. *JAsIs*, *41*(6), 391–407.

Etzkowitz, H. (1998). The norms of entrepreneurial science: cognitive effects of the new university–industry linkages. *Research Policy*, *27*(8), 823–833. doi:10.1016/S0048-7333(98)00093-6

Evans, J. A., & Foster, J. G. (2011). Metaknowledge. *Science*, *331*(6018), 721–725. doi:10.1126/science.1201765

Feller, I. (2006). Multiple actors, multiple settings, multiple criteria: issues in assessing interdisciplinary research. *Research Evaluation*, *15*(1), 5–15. doi:10.3152/147154406781776020

Fleming, L. (2001). Recombinant Uncertainty in Technological Search. *Management Science, 47*(1), 117–132. doi:10.1287/mnsc.47.1.117.10671

Frakes, M., & Wasserman, M. F. (2015). Does the U.S. Patent & Trademark Office Grant Too Many Bad Patents?: Evidence from a Quasi-Experiment. *Stanford University Law Review.*

Granovetter, M. S. (1973). The strength of weak ties. *The American journal of sociology*, *78*(6), 1360–1380.

Hall, B. H., Jaffe, A. B., & Trajtenberg, M. (2001). *The NBER patent citation data file: Lessons, insights and methodological tools*. National Bureau of Economic Research.

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

Haythornthwaite, C. (2006). Learning and knowledge networks in interdisciplinary collaborations. *Journal of the American Society for Information Science and Technology*, *57*(8), 1079–1092. doi:10.1002/asi.20371

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

Jones, B. F. (2009). The Burden of Knowledge and the “Death of the Renaissance Man”: Is Innovation Getting Harder? *The Review of Economic Studies*, *76*(1), 283–317.

Jovanovic, B., & Rob, R. (1990). Long Waves and Short Waves: Growth Through Intensive and Extensive Search. *Econometrica*, *58*(6), 1391–1409.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, *25*(2–3), 259–284. doi:10.1080/01638539809545028

Lattuca, L. R. (2001). *Creating interdisciplinarity: Interdisciplinary research and teaching among college and university faculty*. Vanderbilt university press.

Leahey, E., Beckman, C., & Stanko, T. (2015). Prominent but Less Productive: The Impact of Interdisciplinarity on Scientists’ Research. *arXiv:1510.06802 [cs]*. Retrieved from http://arxiv.org/abs/1510.06802

March, J. G. (1991). Exploration and Exploitation in Organizational Learning. *Organization Science*, *2*(1), 71–87.

Moser, P., Ohmstedt, J., & Rhode, P. W. (2015). *Patent Citations and the Size of the Inventive Step - Evidence from Hybrid Corn* (SSRN Scholarly Paper No. ID 2641659). Rochester, NY. Retrieved from http://papers.ssrn.com/abstract=2641659

Popp, D., Juhl, T., & Johnson, D. K. N. (2004). Time In Purgatory: Examining the Grant Lag for U.S. Patent Applications. *Topics in Economic Analysis & Policy*, *4*(1). doi:10.2202/1538-0653.1329

Porter, A. L., & Rafols, I. (2009). Is science becoming more interdisciplinary? Measuring and mapping six research fields over time. *Scientometrics*, *81*(3), 719–745. doi:10.1007/s11192-008-2197-2

Post, R. C. (1976). Liberalizers" versus “Scientific Men” in the Antebellum Patent Office. *Technology and Culture*, *17*(1), 24–54.

Preschitschek, N., Niemann, H., Leker, J., & Moehrle, M. G. (2013). Anticipating industry convergence: semantic analyses vs IPC co-classification analyses of patents. *Foresight*, *15*(6), 446–464. doi:10.1108/FS-10-2012-0075

Rafols, I., Leydesdorff, L., O’Hare, A., Nightingale, P., & Stirling, A. (2012). How journal rankings can suppress interdisciplinary research: A comparison between Innovation Studies and Business & Management. *Research Policy*, *41*(7), 1262–1282. doi:10.1016/j.respol.2012.03.015

Rosell, C., & Agrawal, A. (2009). Have university knowledge flows narrowed?: Evidence from patent data. *Research Policy*, *38*(1), 1–13. doi:http://dx.doi.org/10.1016/j.respol.2008.07.014

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

Sternitzke, C. (2009). Patents and publications as sources of novel and inventive knowledge. *Scientometrics*, *79*(3), 551–561.

Stuart, T. E., & Podolny, J. M. (1996). Local search and the evolution of technological capabilities. *Strategic Management Journal*, *17*(S1), 21–38.

Sullivan, D., & Lefkowitz, E. (2017). *Patent Quality Chat - Examination Time Analysis.*

Tijssen, R. J. W. (1992). A quantitative assessment of interdisciplinary structures in science and technology: Co-classification analysis of energy research. *Research Policy*, *21*(1), 27–44.

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

Tushman, M. L. (1977). Special boundary roles in the innovation process. *Administrative Science Quarterly*, *22*(4).

Tushman, M. L., & Scanlan, T. J. (1981). Boundary Spanning Individuals: Their Role in Information Transfer and Their Antecedents. *Academy of Management Journal*, *24*(2), 289–305. doi:10.2307/255842

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

Whalen, R. (2015). Second-Order Obviousness: How Information and Communication Technologies Make Inventions More Obvious and Why the Law Should Care. *Journal of the Patent & Trademark Office Society*, *97*(4), 597.

Whalen, R., Huang, Y., Tanis, C., Sawant, A., Uzzi, B., & Contractor, N. (2016). Citation Distance: Measuring Changes in Scientific Search Strategies. In *Proceedings of The 3rd ACM WWW Workshop on Big Scholarly Data: Towards the Web of Scholars*. Presented at the ACM WWW, Montreal.

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

Yegros-Yegros, A., Rafols, I., & D'Este, P. (2015). Does Interdisciplinary Research Lead to Higher Citation Impact? The Different Effect of Proximal and Distal Interdisciplinarity. *PLoS ONE*, *10*(8), e0135095. doi:10.1371/journal.pone.0135095

Youn, H., Strumsky, D., Bettencourt, L. M. A., & Lobo, J. (2015). Invention as a combinatorial process: evidence from US patents. *Journal of The Royal Society Interface*, *12*(106). doi:10.1098/rsif.2015.0272